Review

Frequency-Informed Vision and Learning: A Survey

Lei Zhang *, Tianyu Zhang, Xiaowei Fu, Fuxiang Huang, Wenguan Wang, David Zhang

Posted Date: 11 May 2026

doi: 10.20944/preprints202605.0650.v1

Keywords: frequency-principled deep learning; fourier domain; frequency transforms; computer vision; deep neural networks

 doi:10.20944/preprints202605.0650.v1

Review

Frequency-Informed Vision and Learning: A Survey

Lei Zhang [1,*], Tianyu Zhang [1], Xiaowei Fu [1], Fuxiang Huang [2], Wenguan Wang [3] and David Zhang [4]

[1] School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China
[2] School of Data Science, Lingnan University, Hong Kong
[3] College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[4] School of Science and Engineering, Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China
* Correspondence: leizhang@cqu.edu.cn

Abstract

Frequency, as a physical quantity that describes the rate at which periodic events occur, is a crucial perspective and component, and can help observe and recognize the world via versatile frequency transforms. Initially, built on Fourier analysis theory, it played an important role primarily in the field of signal processing, and has gradually become an indispensable part of deep learning to solve complex problems. Deep learning in the frequency domain (a.k.a. Fourier domain), which we called **F**requency-principled **D**eep **L**earning (FDL), has been extensively employed in a wide range of scenarios owing to its compelling advantages, such as global receptive field, high computational efficiency, inherent data decomposition and explainability. Deep neural networks also exhibit certain properties from a frequency-domain perspective, which provides valuable insights for powerful model design and refinement. Despite growing attention to frequency-domain approaches in deep learning and computer vision, the absence of a systematic synthesis makes it difficult to grasp the current landscape, identify the core methodologies, perceive the challenge, and chart a course for future research. Moreover, a comprehensive explanation for why introducing frequency-domain methods contributes to problem-solving is still lacking. This survey aims to provide a comprehensive and structured overview of frequency-principled vision and learning to address this gap. Unlike previous reviews that may focus on isolated aspects, our work seeks to connect and systematize the field through a unified taxonomy. Specifically, we conduct a systematic survey and analysis of existing literature from multiple perspectives: frequency principle (theory), implementations (algorithms), applications, challenges and future frontiers of FDL across various tasks.

Keywords: frequency-principled deep learning; fourier domain; frequency transforms; computer vision; deep neural networks

1. Introduction

Deep learning has triggered a revolutionary transformation in numerous fields such as computer vision, natural language processing, and signal processing, thanks to its powerful capability of feature learning. From convolution neural networks (CNNs) [61] overcoming the limitations of traditional handcrafted feature extraction in image classification tasks, to the Transformer [130] demonstrating exceptional ability in modeling long-range dependencies in sequential data, deep learning models have achieved remarkable success in analyzing and modeling spatiotemporal-domain data. These achievements have propelled artificial intelligence technology from theoretical research to widespread practical applications. However, with the continuous expansion of application scenarios and the constant growth of data scale, traditional deep learning methods have gradually exposed a series of oblivious issues that significantly constrain their performance improvement and efficiency optimization in complex tasks. For instance, Transformer often exhibits substantial computational complexity when processing large-scale data, such as high-resolution images and long-sequence signals. Convolutional

 doi:10.20944/preprints202605.0650.v1

Neural Networks would be limited in scenarios requiring global modeling due to their local receptive field, and images generated by GANs [107] often suffer from artifacts or detail loss. To address these challenges faced by traditional spatiotemporal-domain deep learning, researchers have begun to turn their attention to frequency-domain analysis.

As an important dimension for data representation, the frequency analysis technique can decompose complex signals in the spatiotemporal domain into a superposition of components with different frequencies. This characteristic enables a more intuitive revelation of the inherent periodicity, global correlations, and frequency-domain feature patterns embedded in data. Against this backdrop, Frequency-principled Deep Learning (FDL) has emerged as a pivotal driving force behind advancements in fields such as computer vision and time series analysis. Specifically, FDL is a type of machine learning paradigm that takes frequency analysis theories such as Fourier transform as the core principle, and deeply integrates deep learning models with frequency-domain characteristics. Its core idea is to break through the limitation of traditional deep learning that only relies on local spatial/temporal features. By mapping data (e.g., images, videos, spatiotemporal sequences) or model parameters to the frequency domain, it leverages the inherent properties of frequency-domain signals such as global correlation and sparsity to achieve more efficient feature extraction, modeling, and optimization. This paradigm can not only address some pain points of traditional methods, including low efficiency [6,46,68,91,129,146], weak robustness [32,67,110,142,184,196], and difficulty in global modeling [12,51,100], but also provide a brand-new perspective for the processing of dynamic and noisy data [34,46,151], acting as a crucial bridge connecting classical signal processing and modern deep learning.

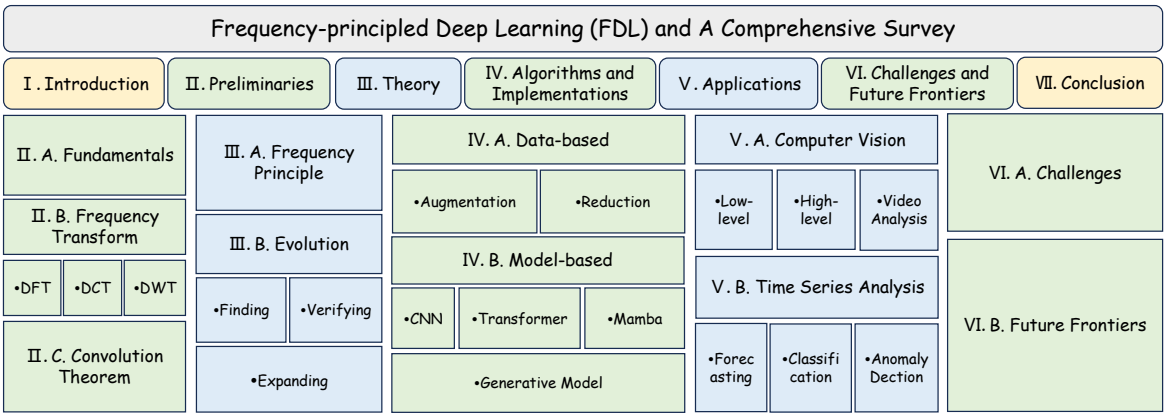

Figure 1. An overview of frequency-principled deep learning (FDL) and the architecture of this article.

The rapid development of FDL is primarily attributed to the continuous breakthroughs in theory and techniques.

On the theoretical front, the refinement of advanced frequency-domain representation theories and the ongoing evolution of the Frequency Principle (F-Principle) have laid a solid mathematical foundation for the intrinsic analysis and feature extraction of signals. Traditional multi-resolution representation methods such as Fourier analysis and wavelet transform have become increasingly sophisticated, while emerging theories including Fourier neural operators [34,70,149] and the F-Principle [155–158,191] have demonstrated tremendous application potential. These theoretical breakthroughs provide valuable guidance for the design of advanced deep learning algorithms and network architectures.

On the technical front, the emergence of flexible integration frameworks has made the combination of frequency analysis and deep networks more systematic and efficient. Such frameworks have propelled the practical implementation and performance enhancement of FDL mainly from two dimensions: data-based and model-based. On the data side, frequency-domain methods are widely adopted for data augmentation [135,151,152,164]. By perturbing or reconstructing frequency components, these methods effectively improve the generalization capability of models. Meanwhile, frequency-principled

 doi:10.20944/preprints202605.0650.v1

representation enables efficient data dimensionality reduction and compression, which cuts down computational and storage costs while accelerating the training and inference process [32,63,144,165]. On the model side, frequency-principled mechanisms are deeply integrated into various mainstream network architectures. For convolutional neural networks (CNNs), techniques such as fast Fourier convolution [12] and spectral pooling [77,110,144,165] enhance the efficiency of feature extraction and expand the receptive field. For Transformers, frequency-domain attention mechanisms can model long-range dependencies and reduce computational complexity [60,96,197]. In addition, generative models leverage frequency-domain constraints to improve the quality and authenticity of generated images [49,94,101,161]. State space models such as Mamba have begun to explore the advantages of frequency-domain representation in temporal modeling [65,120,148,198,202], further broadening the application boundaries of FDL.

The dual evolution of theories and technologies has greatly propelled the application of FDL across a wide range of task scenarios, including image restoration and enhancement [16,29,53,164], object detection [92,135,151,198], segmentation [79,80,114,164], time series analysis [145,160,169,196,197], etc. FDL provides an innovative approach to address the core bottlenecks of traditional deep learning, and has gradually become an important research branch in the field of deep learning. While several related reviews [155,167] have been published to date, a comprehensive overview of Frequency-principled Deep Learning remains still lacking.

This survey aims to systematically delineate the theoretical foundations, technical frameworks, and application scenarios in the field of Frequency-principled Deep Learning, identify common patterns across relevant studies, and sort out key challenges as well as future prospects for the field. Specifically, the main contributions of this survey are summarized as follows: i) We formally define the "Frequency-principled Deep Learning (FDL)" paradigm for the first time, and conduct a systematic survey (covering theories, algorithm implementations, application, challenges and frontiers) which fills the gap of lacking a comprehensive review in this field. ii) We propose a novel dual-level taxonomy based on the stage at which frequency-based methods operate, revealing the diverse implementation strategies of FDL from both data and model perspectives; iii) We identify the key challenges faced by FDL and prospectively point out several promising future development directions.

The overall architecture of this article is shown in Figure 1. Section 2 mainly elaborates on the mathematical foundations of FDL, including the mathematical principles of frequency-domain transforms such as DFT, DCT, DWT as well as the convolution theorem. Section 3 analyzes the importance and necessity of frequency analysis for deep learning models from a theoretical perspective, and summarizes the development history of Frequency Principle. Section 4 provides a detailed analysis and summary of the implementation of FDL paradigms from two aspects: data-based and model-based. Section 5 introduces the applications of FDL in the fields such as computer vision and time series analysis over the past decade, and compares and analyzes the experimental results for several typical tasks. Section 6 discusses the current challenges and future frontiers of FDL. Section 7 summarizes the entire survey.

2. Background and Preliminaries

2.1. Fundamentals of Frequency Domain Analysis

Frequency domain analysis originates from the classical signal processing, i.e., transforming signals from intuitive spatial/temporal domains to abstract frequency domains and revealing inherent imperceptible patterns and regularity hidden in data. This fundamental perspective stems from early breakthroughs in mathematical analysis, such as Fourier's pioneering work [103] on signal decomposition, which laid the groundwork for understanding that any complex signal can be expressed as a superposition of sinusoidal components with distinct frequencies. This insight revolutionized signal processing by shifting the focus from "how signals change over space/time" to "what frequency components constitute signals", formulating the core logic underlying all subsequent frequency-based techniques.

 doi:10.20944/preprints202605.0650.v1

The unique advantages of frequency domain analysis including global receptive field, efficient feature disentanglement, and explicit pattern recognition—are inherently derived from this core idea. Unlike spatial/temporal representations that focus on local or sequential variations, frequency-domain transformation decomposes signals into orthogonal frequency components, enabling direct capture of global correlations, periodicity, and redundancy. For example, low-frequency components often encode overall structures (e.g., image contours, time-series trends), while high-frequency components carry details or noise. This natural separation allows for targeted processing (e.g., denoising by suppressing high frequencies), while that is difficult to achieve in spatial/temporal domains.

To materialize this fundamental idea in practical applications, several key frequency transforms were developed, each tailored to specific scenarios while adhering to the unified principle of frequency decomposition:

Discrete Fourier Transform (DFT) [118]: A foundational transform for digital signal processing and a discrete implementation of Fourier's decomposition theory, enabling quantitative analysis of frequency components in signals.

Fast Fourier Transform (FFT) [4]: An optimized algorithm for DFT, that drastically reduces computational complexity, making frequency-domain operations feasible for large-scale data processing.

Discrete Cosine Transform (DCT) [1]: Specialized for real-valued signals (e.g., images, audio), that retains DFT's frequency representation capability while remaining only real numbers, optimizing efficiency in filters.

Discrete Wavelet Transform (DWT) [89]: Addresses the limitation of DFT/DCT in balancing time-frequency resolution, that uses local wavelet basis functions to capture both global trends (low-frequency) and local details (high-frequency), facilitating data compression [165] and denoising [124].

These transforms, rooted in the core idea of frequency decomposition, collectively form the technical foundation of frequency domain analysis. Their application spans computer vision [135,150, 164], time series analysis [169,197], etc., where analyzing signal amplitude (magnitude of frequency components) and phase (positional relationship of components) reveals critical spectral characteristics. For instance, the Scattering Vision Transformer [97] leverages frequency separation to enhance image detail retrieval. Additionally, Fourier analysis provides a key theoretical lens for understanding deep neural networks: numerous spectral bias studies [108,154,156–158,191] show that gradient descent prioritizes low-frequency components during training, highlighting the intrinsic connection between frequency principles and deep learning.

2.2. Frequency Transform

Frequency-domain transformation serves as the foundation of frequency-principled deep learning. It functions as converting signals that are difficult to analyze directly in the spatiotemporal domain—such as images, audio, and physiological signals—into the frequency domain. By revealing the inherent frequency components within the signals, it provides more effective, intuitive and explicit data representations beneficial for subsequent feature extraction and model construction. DFT [118], DCT [1] and DWT [89] are the three most well-known and widely used transforms in FDL. Each possesses distinct mathematical characteristics and application scenarios, collectively formulating the core technical framework for frequency-domain data processing. In this section, we provide a brief introduction to these three frequency-domain transforms.

2.2.1. Discrete Fourier Transform

DFT [118] is an extension of the Fourier transform for discrete signals. Its essence lies in decomposing a discrete spatiotemporal domain signal $x[n]$ of length N (where $n = 0, 1, ..., N-1$) into a linear combination of N complex exponential signals (i.e., trigonometric function) with different frequencies,

 doi:10.20944/preprints202605.0650.v1

thereby explicitly describing the frequency components and phase information of the signal in the frequency domain. DFT is mathematically defined as

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j \cdot 2\pi kn/N} \tag{1}$$

where $k = 0, 1, \ldots, N-1$, $X[k]$ represents the complex-valued frequency-domain signal, with its real part representing the amplitude and its imaginary part representing the phase. $e^{-j \cdot 2\pi kn/N}$ denotes the complex exponential basis function of frequency k/N and j means the imaginary unit. The inverse transform reconstructs the spatiotemporal domain signal $x[n]$ from the frequency-domain signal $X[k]$, ensuring lossless conversion between the two domains.

2.2.2. Discrete Cosine Transform

DCT [1] is a modified frequency-domain transform method for real-valued signals, building upon the foundation of DFT. Since most signals in practical applications such as image and audio are real-valued, the complex output of DFT contains redundancy where the conjugate symmetric portion can be derived as real numbers. By directly employing cosine basis functions instead of complex exponential basis functions, DCT confines the transform results to the real number domain while retaining the global frequency representation capability of DFT. Generally, DCT can be easily deduced by operating DFT on an even function.

DCT has multiple variant definitions (e.g., from DCT-I to DCT-VIII), among which, DCT-II is most widely used in image processing. Mathematically, it is defined as

$$X[k] = \alpha(k) \sum_{n-0}^{N-1} x[n] \cdot \cos\left(\frac{\pi(2n+1)k}{2N}\right) \tag{2}$$

where $k = 0, 1, \ldots, N-1$, $\alpha(k)$ is a normalization coefficient that ensures the orthogonality and defined as:

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{N}} & \text{if } k = 0 \\ \sqrt{\frac{2}{N}} & \text{if } k \neq 0 \end{cases} \tag{3}$$

The cosine basis function $\cos\left(\frac{\pi(2n+1)k}{2N}\right)$ is a real-valued function, and thus $X[k]$ is real and directly corresponds to the amplitude of the signal at frequency k.

2.2.3. Discrete Wavelet Transform

DFT and DCT are based on global basis functions, making them difficult to balance frequency resolution and time/spatial resolution, i.e., they cannot accurately locate where the frequency components of a signal occur. Therefore, DWT with wavelet basis functions (i.e., local basis functions with finite duration and rapid decay characteristics) instead of global trigonometric function is evolved [89]. DWT achieves precise analysis of local time-frequency features of signals, filling the gap of global transforms in capturing local features. DWT employs multi-scale decomposition to decompose a signal into low-frequency approximation components and high-frequency detail components of different resolutions (scales). Here, the low-frequency approximation component reflects the overall trend of the signal and is extracted using the low-frequency portion of the wavelet basis function, while the high-frequency detail component captures localized abrupt changes (e.g., edges or noise) via the high-frequency portion of the wavelet basis function.

Mathematically, DWT realizes multi-scale convolution of the input signal through scaling and translation of the wavelet basis function $\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$. Then, DWT is defined as:

$$W[j,k] = \int_{-\infty}^{+\infty} x(t) \cdot \psi_{j,k}(t) dt \tag{4}$$

 doi:10.20944/preprints202605.0650.v1

where j is the scaling factor, k is the translation factor, and $W[j,k]$ is the wavelet coefficient which reflects the similarity between the signal and the wavelet basis function at scale j and position k. This achieves adaptive analysis with high time resolution for high-frequency signals and high frequency resolution for low-frequency signals.

2.3. Convolution Theorem

Convolution Theorem [116] states that the convolution operation in the temporal-domain is equivalent to pointwise multiplication in the frequency domain. Given a temporal-domain input signal $x(t)$ and a filter $h(t)$, the Convolution Theorem can be mathematically described as:

$$\mathcal{F}(x(t) * h(t)) = \mathcal{F}(x) \cdot \mathcal{F}(h) \tag{5}$$

where $\mathcal{F}(\cdot)$ denotes the Fourier Transform, and $\mathcal{F}(x)$ and $\mathcal{F}(h)$ represent the Fourier transforms of $x(t)$ and $h(t)$, respectively.

In traditional deep learning (i.e., spatial-domain), the convolution operation requires sliding between the input feature map and the convolution kernel. As the input size and the number of convolution kernels increase, the computational complexity rises exponentially, significantly constraining the operational efficiency of the model. The Convolution Theorem enables deep learning models to convert high-complexity spatial convolutions into low-complexity frequency domain multiplications. This transform not only substantially reduces computational load particularly for processing large-scale inputs or large convolution kernels, but also decreases the memory usage of the model.

Building upon this foundational advantage, the Convolution Theorem has been widely applied in the field of deep learning. Early pioneering works, such as accelerating convolutional network training via FFTs [91] and implementing efficient training of convolutional deep belief networks directly in the frequency domain [6], demonstrated its potential for drastically accelerating model training speed. This principle continues to inspire more sophisticated algorithmic designs. A prominent example is the development of frequency-domain attention mechanisms. Representative works such as FSAS [60] and Autoformer [146] transform Query and Key into the frequency domain via FFT, then compute their correlation through an element-wise product operation instead of computing the matrix multiplication in the spatial domain. This design reduces the time complexity of self-attention from $O(N^2)$ to $O(NlogN)$ while maintaining comparable performance.

3. Frequency-Principled Deep Learning: Perspective of Frequency Analysis Theory

3.1. Frequency Analysis and Frequency Principle

The success of deep learning in several fields such as computer vision and natural language processing is closely associated with its complex parameter space and nonlinear characteristics. However, these properties further lead to the "black box" of deep models, which make the training process and generalization mechanisms difficult to interpret. Traditional theories struggle to accommodate the over-parameterized nature of deep models [31]. Frequency analysis, by decomposing signals into different frequency components, reveals the model's learning preferences for features of varying complexities. This frequency perspective not only offers an explanation for the key question: *why over-parameterized models do not overfit easily*, but also provides a quantitative basis for optimizing model robustness and architecture design. This section aims to uncover the profound connections between frequency analysis and deep learning, systematically summarizing and analyzing the following aspects: the discovery and validation of the Frequency Principle (F-Principle), including the frequency-based interpretation of model generalization capability and frequency-domain analysis of robustness, and the application of F-Principle in specific tasks.

The Frequency Principle is a key characteristic of FDL. Its core proposition states that: deep neural networks trained with gradient descent preferentially fit the low-frequency components of the data before progressively learning the high-frequency components. This phenomenon is not confined to specific scenarios but is universally observed across different architectures, datasets, and training

 doi:10.20944/preprints202605.0650.v1

settings. Related studies [3,108,154–158,191] have established a foundation of the Frequency Principle through theoretical derivation and empirical validation.

3.2. Evolution of Frequency Principle

3.2.1. Finding: Spectral Bias during DNN Training

In the field of deep learning, the traditional perspective on generalization suggests that the generalization error can be controlled by model complexity. While more complex models can better fit the training data, they are also more prone to overfitting, leading to a decline in generalization performance. However, in practice, deep neural networks, despite their complex structures and numerous parameters, often exhibit excellent generalization capabilities.

To address this paradox, some studies have attempted to understand the generalization ability of DNNs from the perspective of the training process. Arpit et al. [3] pointed out that during gradient optimization, DNNs prioritize learning the simple patterns of the true data. Based on Fourier analysis theory, Xu et al. [154,158,191] proposed the Frequency Principle, demonstrating that during training, DNNs first rapidly capture the dominant low-frequency components of the data while keeping their own high-frequency components small, and then progressively capture the high-frequency components. This is coined as spectral bias. Rahaman et al. [108] further validated the existence of spectral bias in neural networks during gradient descent training, where low-frequency components are learned more quickly and are more robust to parameter perturbations, and highlighted Fourier analysis as an effective tool for understanding the inherent properties of neural networks.

3.2.2. Verifying: Universality of F-Principle

Xu et al. [156,157] theoretically analyzed the DNN training process based on gradient methods through Fourier analysis, explaining why DNNs with small initializations can achieve good generalization ability, thereby further refining the F-Principle from a theoretical perspective. Then they demonstrated that the Frequency Principle holds for general loss functions, not limited to Mean Squared Error loss [154]. Xu et al. [191] further validated the universality of the Frequency Principle across various scenarios, including MNIST/CIFAR10 datasets and VGG16 network, by designing projection methods and filtering methods. In other words, the F-Principle holds across different DNN architectures (e.g., fully-connected networks and convolutional neural networks), different activation functions (e.g., tanh and ReLU), and different loss functions (e.g., MSE, cross-entropy and variational loss).

3.2.3. Expanding: Deep F-Principle for Faster Training

Built upon the F-Principle, Xu et al. [158] further proposed the Deep Frequency Principle, which pursues an effective target function for a deeper hidden layer bias towards lower frequency during training. Since it has been verified that "*DNNs tends to learn low-frequency functions faster*", if the learning components in deeper networks have an effective target function biased towards lower frequencies, deeper networks are hopeful to finish training in fewer epochs, i.e., faster training can be achieved. An overview of the F-Principle can be referred to as [155].

4. Algorithms and Implementations

4.1. Data-Based

4.1.1. Data Augmentation

With the rapid development of deep learning in the field of computer vision, data augmentation has become a key technique for alleviating data scarcity and enhancing the generalization ability of deep models. Traditional spatial-domain data augmentation methods (such as flipping, cropping, and MixUp) achieve augmentation by applying geometric or pixel-level perturbations to pixel-wise data (e.g., image). However, these methods often suffer from issues such as coarse operation granularity, weak adversarial robustness, semantic distortion, low diversity, etc. By performing selective operations

 doi:10.20944/preprints202605.0650.v1

on different frequency components, frequency-domain data augmentation introduces reasonable perturbations while preserving core semantics of data, thereby overcoming the limitations of spatial-domain augmentation. In recent years, frequency-domain data augmentation has demonstrated broad application prospects in various fields, such as transfer learning, few-shot learning, adversarial example generation, etc. The basic paradigm of frequency-domain data augmentation is depicted in Figure 2.

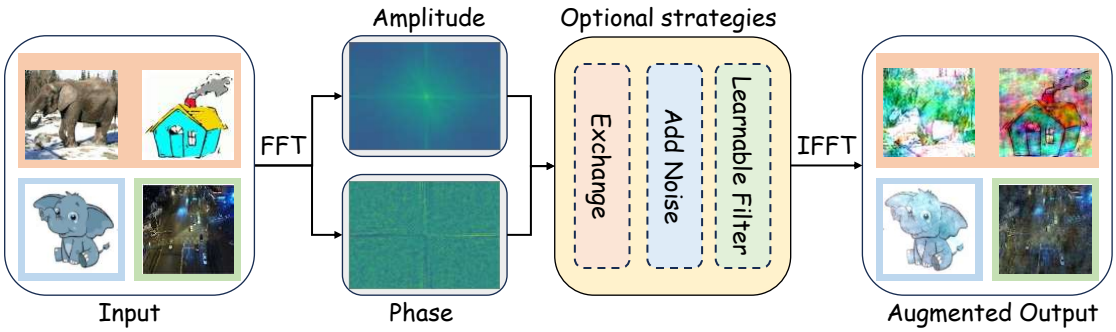

Figure 2. Data augmentation in the frequency domain, where *Exchange* means that swapping the amplitude or phase in frequency-domain of input, *Add Noise* means that adding perturbations to the input amplitude or phase, and *Learnable Filter* means that performing adaptive filtering on the input amplitude and phase to adapt specific scenarios.

Data augmentation is a core technique for enhancing the generalization ability of deep learning models. Training deep learning models typically requires massive amounts of data. However, collecting new data spends a great deal of time and effort, while it is more efficient if one can augment more data from existing data in many scenarios. The unique decoupling capability and global representation of frequency-domain transform hold a great potential in the field of data augmentation. Yang et al. [164] pioneered a simple unsupervised domain adaptation method. They observed that significant changes in low-level spectral components (e.g., amplitude) do not affect the perception of high-level semantic information, and thus proposed to achieve data augmentation by swapping the low-frequency amplitude spectra between source and target domain images. Inspired by Yang et al.'s work, Xu et al. [152] extended this Fourier-based data augmentation method to the domain generalization scenario where no target domain data is available. FACT [152] performs linear interpolation on the amplitude spectrum of source domain images to generate augmented images. By perturbing the amplitude information while preserving the phase information, it forces the model to learn high-level semantics from the phase spectrum. FDAG proposed by Wang et al. [133] introduces noise in both amplitude and phase to achieve more comprehensive perturbations in the frequency domain.

Similar to the aforementioned works, by decomposing images into multiple frequency components and only perturbing those components with low semantic information, FAA [43] can generate adversarial examples with large perturbation magnitudes yet minimal semantic modifications. FSDR [42] leverages DCT to decompose images into domain-invariant frequency components (DIFs) and domain-variant frequency components (DVFs). It only randomizes DVFs while keeping DIFs unchanged, thereby minimizing the impact on the semantic structure (content) distortion of images. Wang et al. [135] suggested separating domain-invariant and domain-specific spectral components from the amplitude spectrum while preserving the phase spectrum to ensure the integrity of the image structure. Xu et al. [151] argued that extremely high and low frequency components contain more non-causal factors and proposed a non-causal image augmentation method that preserves causal features unchanged and randomizes non-causal frequency components following a Gaussian distribution.

4.1.2. Feature Dimensionality Reduction

In the field of deep learning, practical tasks often face the curse of dimensionality, which means that excessively high feature dimensionality may lead to a surge in computational complexity and

 doi:10.20944/preprints202605.0650.v1

an increased risk of model overfitting. Feature reduction refers to reducing the dimensionality of the feature space while preserving key information, and it serves as a critical technique for model optimization. In natural images, most of the critical structural information (e.g., contours, shapes and backgrounds) is concentrated in low-frequency components, while high-frequency components usually contain noise and fine-grained texture details. Based on this observation, frequency-domain downsampling has emerged and gradually matured. Thanks to this unique energy compaction property, frequency-domain deep learning naturally exhibits significant advantages in feature dimensionality reduction.

Levinskis et al. [63] were the first to integrate DWT into CNNs, using only approximation coefficients to substantially reduce dimensionality without losing essential features. Williams et al. [144] proposed wavelet pooling to replace the traditional downsampling operation. Li et al. [67] improved the model's noise robustness through similar operations. Yao et al. [165] employed DWT to perform lossless downsampling on the keys and values of self-attention networks, which alleviates the high complexity of Transformers and reduces the information loss in traditional downsampling. In addition, converting feature maps to frequency domain via DFT and retaining low-frequency components can also achieve valid feature dimensionality reduction [21,32,110]. In a word, frequency-based deep learning aggregates information scattered in the spatial domain, thereby making it possible to retain only key frequency components. This holds extremely broad application prospects in handling tasks with complex inputs such as ultra-high-resolution images and real-time video streams.

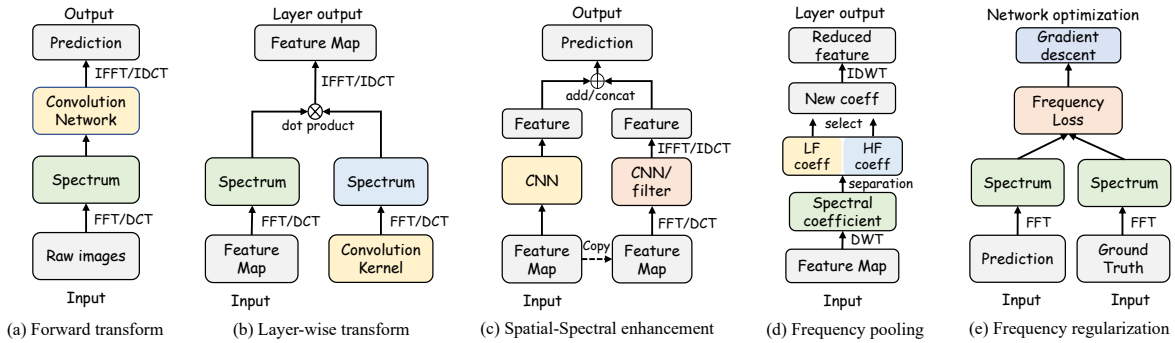

Figure 3. Versatile strategies for exploring frequency transform in convolutional neural networks. The first 4 items denote different architecture for FDL upon CNN, while the final one denotes the frequency based regularization loss.

4.2. Model-Based

This section systematically outlines the research roadmap and context of model-based methods under frequency principle with four mainstream network architectures, such as CNNs, Transformers, Mamba and Generative models.

4.2.1. CNN-Based

Integrating frequency transforms into CNNs offers several benefits: enhanced feature extraction through multi-resolution analysis, improved computational efficiency through optimized frequency domain operations, and robust performance with frequency regularization. These advancements facilitate the development of FDL techniques. Based on differences in integration locations and target components, we categorize the frequency-domain enhancement strategies commonly used in CNNs into five paradigms, including forward transform on raw images, layer-wise transform, spatial-spectral parallel enhancement, frequency pooling and regularization, as summarized in Figure 3.

Forward transform. The most straightforward way to integrate frequency-based module into CNNs is to place the frequency transform at the network's input layer as a data preprocessing step. As shown in Figure 3a, the raw image is first converted into a spectrogram via FFT or DCT, and then the spectrum is then fed into the subsequent convolutional layers for processing. Compared with traditional spatial-domain deep learning, this paradigm offers two primary advantages: low computational complexity

 doi:10.20944/preprints202605.0650.v1

and the decoupling property of frequency-domain features. The Convolution Theorem states that complex spatial convolution operations can be simplified to element-wise multiplication of frequency-domain coefficients. Early works [6,91] have fully verified its enormous potential in reducing training and inference time costs, with specific details elaborated in Section 2.3. In addition, the raw images to be processed can be fed into CNNs in the form of frequency-domain coefficients. For instance, Gueguen et al. [33] proposed a ResNet architecture that operates directly on DCT coefficients instead of pixels. This method can bypass redundant JPEG decoding and RGB conversion process, enabling more efficient training and inference. The inherent decoupling property of frequency-domain features enables CNNs to filter out noise and redundant information in the preprocessing stage, thereby directly focusing on the task-relevant frequency-domain components. And this also means that FDL can achieve more diverse and effective data augmentation, with details provided in Section 4.1.

Layer-wise transform. In CNNs, the frequency transform is applicable not only to input data but also to the model parameters. This paradigm applies frequency transform to the internal components of the model (e.g., convolutional kernel weights), aiming to reduce frequency-domain redundancy inherent in parameters. Convolutional kernels are generally smooth and exhibit spatial-domain redundancy, which is typically manifested as high energy concentration in the DCT domain. Inspired by this observation, researchers have achieved efficient model compression by transforming filter weights into the DCT domain and discarding low-energy coefficients [138], or performing dynamic pruning on frequency-domain parameters [84]. Furthermore, FreshNets [10] leverages hash functions to group frequency-domain model parameters into hash buckets, thereby achieving parameter sharing. Such methods demonstrate distinct advantages in scenarios that pursue extreme model compactness, yet they require careful design of compression thresholds to prevent drastic performance degradation caused by the loss of critical high-frequency parameters. Another category of works directly define and optimize filters in the frequency domain [7], or replace entire convolutional layers with DCT-based "harmonic blocks" [128]. This is equivalent to imposing an implicit spectral smoothing constraint on model learning, making the model easily converge to structured solutions.

Spatial-Spectral enhancement. Convolutional neural networks naturally excel at capturing local spatial features due to their local receptive field property, yet this also introduces the limitation of insufficient modeling of long-range dependencies. This contradiction has motivated researchers [45, 132,166,172] to incorporate spectral branch into CNNs, leveraging the inherent ability of the frequency domain to represent global information. By integrating the local detail information from the spatial branch and the global structural information from the spectral branch, frequency-based CNNs can obtain higher-quality feature representation.

Frequency Pooling. The downsampling operation of the pooling layer can also be redesigned via frequency transformation to achieve downsampling with more sufficient information retention, as shown in Figure 3d. Fujieda et al. [24,25] claimed that traditional CNNs represent a limited form of multi-resolution analysis. By repeatedly performing convolution and pooling operations on input data, they are essentially equivalent to utilizing only the low-frequency components of multi-resolution analysis. The wavelet transform can decompose an image into low-frequency coefficients and high-frequency coefficients at different levels, thereby capturing global structural information and local textural details. After selecting and reallocating these frequency coefficients, frequency pooling can enable more rational downsampling. For instance, Williams et al. [144] first proposed the concept of wavelet pooling, where they suggested using wavelet transform to perform two-level decomposition on feature maps, discarding the high-frequency features of the first-level subbands while retaining those of the second-level subbands, thereby achieving feature dimensionality reduction. MWCNN [78] replaces the pooling operations of traditional U-Net with the DWT to obtain a larger receptive field. However, Finder et al. [22] argued that these methods [2,25,78] are highly customized architectures and cannot be adopted in other CNN frameworks. They further proposed a new layer named WTConv that can effectively expand the receptive field of CNNs, serving as a plugged-played replacement

 doi:10.20944/preprints202605.0650.v1

component for deep convolutional layers. In addition, Grabinski et al. [32] achieved aliasing-free downsampling by cropping components above the Nyquist frequency in the Fourier frequency domain.

Frequency regularization. Integrating spectral constraints as optimization objectives into the deep models has been widely applied in various FDL models [13,29,145,183]. Unlike traditional methods that use the frequency domain as a feature extraction space, this paradigm regards the frequency domain as a diagnostic and constraint space for model behavior. By introducing penalty terms based on spectral characteristics into the loss function, it explicitly guides the model to learn the desired frequency response patterns. For instance, Gao et al. [29] proposed a frequency-domain contrastive regularization term, which forces the model to reduce the distance between de-rained images and clear images (positive samples) in frequency domain, while increasing the distance between de-rained images and various rain-streaked images (negative samples).

4.2.2. Transformer-Based

The self-attention mechanism of Transformers [130] achieves global information fusion by calculating the correlations between all token pairs. However, its computational complexity of $O(N^2)$ and memory consumption grow quadratically with the sequence length N, making it difficult to apply for high-resolution image and video processing.

One approach to this problem is to replace the self-attention layer with an MLP-based mixer layer [75,126,127]. Inspired by this, several works [34,46,62,109,123,178] attempted to improve the self-attention mechanism using spectral mixing techniques. FNet [62] first replaces self-attention sub-layers in Transformer encoders with standard and parameter-free Fourier Transform, significantly reducing the computational load while maintaining performance. Inspired by the convolution theorem, GFNet [109] proposes a global filter to replace the self-attention sub-layer. By performing element-wise multiplication between frequency-domain features and a learnable global filter, GFNet achieves efficient token mixing. Following this work, Huang et al. [46] also claimed that adaptive frequency filters can serve as effective global token mixers. Based on the MetaFormer architecture [173], Tatsunami et al. [123] designed a novel token mixer called Dynamic Filter. Guibas et al. [34] framed token mixing as an operator-learning task that maps continuous functions into an infinite-dimensional space, and designed the Adaptive Fourier Neural Operator to serve as a token mixer. SPANet [178] achieves the balance between high- and low-frequency components through frequency-domain mask filtering, and verified the hypothesis that balancing high-frequency and low-frequency representations can improve model performance. The basic paradigm of frequency-based transformer is depicted in Figure 4.

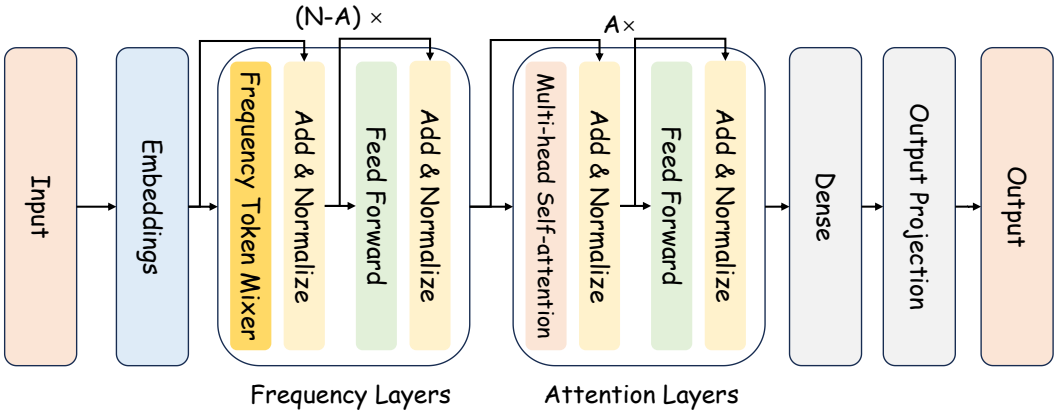

Figure 4. The architecture of frequency-based Transformer, where N denotes the number of Transformer encoder blocks, A denotes the number of attention layers, and $N - A$ means the number of frequency layers referring to blocks that adopt frequency-based spectral mixing strategies.

Another way is to improve the attention mechanism from the perspective of frequency analysis. Based on the convolution theorem, Kong et al. [60] converted the matrix multiplication of queries and keys in the spatial domain to element-wise product in the frequency domain, reducing the spatial

 doi:10.20944/preprints202605.0650.v1

and temporal complexities to $O(N)$ and $O(NlogN)$, respectively. The high-frequency components of an image are rich in local details, while the low-frequency components focus on the global structure. However, the multi-head self-attention layer neglects the characteristics of different frequency components. From this perspective, Pan et al. [96] proposed the HiLo attention mechanism, which splits the attention heads into two groups to separately interpret the high/low-frequency patterns in the attention layer. To address the low-rank issue of the attention matrix caused by frequency-domain sparsity, Yue et al. [177] introduced an additional learnable matrix into the original attention matrix, followed by the row-wise l_1 normalization. SpectFormer [98] achieves outstanding performance by fusing FFT-based spectral layers and multi-head attention layers, and pointed out that both spectral layers and attention layers are essential for transformer.

4.2.3. Mamba-Based

Frequency-domain Mamba is a novel deep learning paradigm that integrates the efficient long-sequence modeling capability of state space models (SSM) with frequency-domain analysis techniques. Mamba achieves long-sequence modeling with linear complexity through a selective scan mechanism, breaking through the quadratic complexity limitation of Transformer-like models [82]. Frequency-domain transforms decompose images into complementary frequency components, providing intuitive feature representations, therefore their combination is beneficial to performance and computational efficiency in complex visual tasks.

The DWT possesses excellent multi-resolution decomposition capability, enabling it to split an image into low-frequency and high-frequency sub-bands. This facilitates Frequency-domain Mamba in designing differentiated modules for extracting global structural information of low-frequency components and detailed information of high-frequency components. Wave-Mamba [200] focuses on the restoration of low-frequency global information in the Low-Frequency State Space Block, and then corrects high-frequency details using the enhanced low-frequency information in the High-Frequency Enhancement Block. WaveMamba [198] first performs low-frequency feature fusion on RGB images and infrared images through channel swapping and gated attention. Meanwhile, it adopts an "absolute maximum" strategy to enhance high-frequency components, and finally reconstructs features via the inverse DWT to reduce information loss. Tan et al. [120] found that swapping wavelet low-frequency components can improve brightness more effectively than swapping Fourier amplitude components, and swapping Fourier phase components can make up for the deficiency of wavelet high-frequency components in detail restoration. Based on this frequency prior, they proposed an Encoder-Latent-Decoder structure: the Wavelet-based Mamba Block is adopted in the Encoder and Decoder for global brightness adjustment, while FFT is adopted in the Latent layer for local detail enhancement.

The global modeling capability of Fourier transform has demonstrated unique advantages in various tasks, and how to use Mamba in the Fourier domain is becoming a hot topic. Li et al. [65] proposed a novel FourierMamba. By applying different Fourier scanning strategies in both the spatial and channel dimensions, FourierMamba successfully models the ordered dependencies between different frequencies in the frequency domain. Xiao et al. [148] designed a frequency selection module based on FFT to identify and select the most informative frequency components as additional cues for Mamba. Zou et al. [202] proposed a three-branch architecture named FreqMamba, which performs 2D Mamba scanning in the original spatial dimension to capture local details and spatial correlations, conducts scanning from low to high frequencies in the frequency band dimension to capture dependencies between different frequencies, and applies convolution in the Fourier domain to model the global degradation patterns of images, thereby realizing collaborative modeling across the spatial, frequency band, and Fourier domains. Mamba's superior performance in modeling complex relationships makes it a compelling choice to synergize frequency domain techniques.

4.2.4. Generative Model-Based

Generative models have achieved remarkable progress in computer vision tasks such as image synthesis and style transfer. However, traditional spatial-domain generative models generally suffer

 doi:10.20944/preprints202605.0650.v1

from the problem of spectral bias [49,56,112,121], leading to problems like poor quality in high-frequency detail generation and frequent occurrence of grid artifacts. By migrating the generation and discrimination processes to the frequency domain, frequency-domain generative models effectively address issues such as high-frequency detail loss and grid artifacts, while achieving breakthroughs in both generation efficiency and quality. This section mainly focuses on innovative works of GANs and diffusion models in frequency domain.

Recent works on frequency-domain GANs mainly focus on generators and discriminators, and the basic structure is shown in Figure 5. Liu et al. [83] designed a wavelet-based discriminator that employs wavelet packet transform (WPT) to extract multi-scale texture features. Compared with convolutional layers for extracting multi-scale features, WPT can significantly reduce computational costs. In addition to the traditional spatial discriminator, Jung et al. [54] also introduced the second discriminator in the frequency domain, which acts directly on the power spectrum of real and generated images, enabling the generator to learn real distributions of the spatial and frequency domains simultaneously. Chen et al. [11] embedded a frequency-aware classifier into the original discriminator, enabling the single discriminator to assess the authenticity of inputs in the spatial and frequency domains simultaneously. Building on the StyleGAN2 [55], SWAGAN [26] integrates wavelet transform into the generator and discriminator: the generator directly predicts coefficients in the wavelet domain, while the discriminator analyzes both the RGB space of images and their entire wavelet decomposition simultaneously. By introducing the wavelet discriminator from SWAGAN, StyleSwin [179] effectively suppresses the block artifacts in high-resolution image generation.

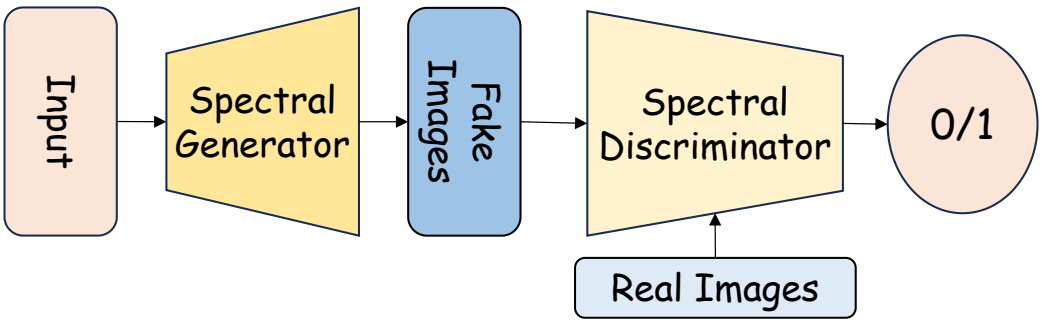

Figure 5. The architecture of frequency-based GANs. The main difference between frequency-based GANs and traditional GANs lies in the generator and discriminator. The Spectral Generator usually introduces additional frequency-domain information. The Spectral Discriminator not only distinguishes real and fake samples in the spatial domain, but also requires the generated images to be consistent with real images in terms of frequency-domain distribution.

It has been confirmed that convolutional GANs inherently suffer from spectral bias, with high-frequency information being more prone to lose. Khayatkhoei et al. [56] argued that this low-frequency preference is controllable and proposed the Frequency Shifted Generator, which transfers the model's low-frequency generation capability to the high-frequency via a frequency shift operator. Jiang et al. [49] proposed an innovative focal frequency loss, enabling the model to pay more attention to the frequencies difficult to synthesize. To address the overfitting problem of discriminator under limited data, Yang et al. [141,161] used DWT to decompose the intermediate features into different frequency components, thereby fully leveraging the frequency information, especially the high-frequency information.

Diffusion models generate high-quality images through progressive denoising and demonstrate superior performance compared to GANs in various cases. However, their training and inference speeds are extremely slow, which greatly limits their applications in practical scenarios. Phung et al. [101] proposed a DWT-based diffusion model called WaveDiff, which performs denoising and sampling in the wavelet domain. By means of wavelet decomposition, WaveDiff extracts low-frequency and high-frequency components from both the image and feature levels. As the spatial dimension is

 doi:10.20944/preprints202605.0650.v1

reduced to 1/4, the computational complexity is significantly decreased. Yang et al. [163] specifically designed a Spectral Diffusion framework for lightweight diffusion probabilistic models, aiming to enhance the capability of lightweight models in generating high-frequency details. DCTDiff [94] trains the diffusion model directly on DCT coefficients, avoiding the high-dimensional redundancy in the pixel space as well as the complexity and training costs of latent diffusion models. Qian et al. [104] proposed a sampling enhancement method for diffusion models. By introducing moving average sampling in the frequency domain, the stability of denoising process and the quality of generated images are promoted.

Table 1. Related works and key contributions on low-level vision with FDL paradigms.

Category	Method	Key Points
Image Restoration	SPHformer [52], WF-Diff [185], SFNet [16], LaMa [119], LaMa-UFFC [14], MWCNN [77], PW-FNet [51], Fourmer [195], D^3 [140], Wave-fill [174]	FFT Loss [52] Frequency Prior [52,185,195], Attention Mechanism [185], Spatial-Frequency Fusion [185], Band-specific Restoration [174], Frequency Dynamic Selection [16], FFC [119], UFFC [14], Wavelet Pooling [77], WT-FFT [51], Frequency Transformer [51]
Image Deblurring	SDWNet [201], MIMO-UNet [13], FourierDiff [86], MLWNet [30], FFTformer [60], DeepRFT [90]	Frequency Loss [13,90], Fourier Prior [86], Learnable DWT [30], Frequency Transformer [60], Attention Mechanism [60], FFT-Conv [90]
Image Dehazing	FrDiff [74], FSDGN [172], DW-GAN [23]	Frequency Priors [74,172], Attention Mechanism [74], Frequency Loss [172], Spatial-frequency Interaction [172]
Image Denoising	FDK [57], SFANet [36], MWDCNN [124], FGDNet [115], EFF-Net [47]	Spectral Discriminator [57], Frequency Loss [57,115], Attention Mechanism [36,47], Frequency Decomposition [47,115]
Image Demoireing	WDNet [76], MBCNN [187,189]	Wavelet Loss [76], Wavelet-Based Dual-Branch Network [76], Frequency Domain Priors [187,189], Learnable bandpass filter [187,189]
Image Deraining	FreqMamba [202], FourierMamba [65], FPNet [38], FADformer [29]	Frequency Mamba [65,202], Frequency Loss [38,65,202], Dual-dimensional Fourier scanning [65], Fourier Prior [38], Frequency-aware Transformer [29], Contrastive Regularization [29]

5. Application and Practices

5.1. Low-Level Vision

5.1.1. Image Restoration and Denoising

Image restoration aims to recover high-quality visual content from degraded images caused by factors such as noise, blur and adverse weather, etc. However, traditional spatial-domain methods such as CNN-based and Transformer-based face inherent limitations. The former struggle to capture long-range dependencies due to limited receptive fields, while the latter suffer from quadratic complexity. These approaches can not fully leverage the frequency domain differences between clean and degraded images.

Some researchers [51,52,60] would like to accomplish image restoration tasks in the frequency domain to reduce computational complexity. [60] develops an efficient Frequency-domain Self-Attention Solver based on the convolution theorem, which replaces spatial matrix multiplication with element-wise products in the frequency domain. [52] integrates FFT mechanism into the Transformer architecture to address the high computational complexity of self-attention. [51] effectively reduces computational complexity while maintaining global restoration capabilities by integrating wavelet

 doi:10.20944/preprints202605.0650.v1

decomposition with FFT. The frequency domain inherently possesses a global perspective, where each frequency component correlates with all spatial pixels. [14,119] introduces Fast Fourier Convolution (FFC) [12] into image restoration, which transforms features into the frequency domain and performs Channel-wise Fully Connected operations, thereby achieving global receptive fields at a lower computational cost.

Numerous works [38,44,52,86,171,185,188,195] leverage frequency domain priors for image restoration. Frequency transform can separate degradation factors from content components. Through Fourier analysis, [185,195] indicate that the amplitude spectrum primarily contains degradation information, while the phase spectrum preserves structural information. Furthermore, [52] found that different types of degradation manifest distinctly across frequency bands. By strategically swapping frequency components, it becomes possible to achieve "degrading clean images" and "restoring degraded images". FPNet [38] and FECNet [44] employ a two-stage network to process amplitude and phase spectrum respectively, achieving simultaneous image restoration and structural refinement. FourierDiff [86] embeds Fourier priors into a pre-trained diffusion model, guiding the reverse diffusion sampling process through amplitude-phase decomposition to facilitate image deblurring. MBCNN [188] learns frequency domain priors of moiré patterns through multi-block learnable bandpass filters for moiré pattern removal. [171] proposes a dual-branch network FSDGN that processes frequency-domain and spatial-domain information in two respective branches, utilizing the residual of amplitude spectrum to guide the restoration of phase spectrum. An overview of image restoration methods under the FDL paradigm is presented in Table 1.

5.1.2. Image Compression and Image Super-Resolution

Image Compression. In the field of image compression, since the energy of an image is concentrated in the low-frequency components, the classic JPEG [131] algorithm achieves lossy compression by employing a quantization matrix in the DCT domain to reduce the precision of high-frequency components. In deep learning-based compression, the application of frequency analysis theory goes beyond explicit DCT, becoming more flexibly and deeply embedded within the network architecture. Guo et al.'s DDCN [35] constructs a parallel architecture operating in both the DCT domain and the pixel domain. The DCT domain branch exploits redundancy by analyzing the distribution of frequency domain coefficients, while the pixel domain branch incorporates frequency domain priors to help detail recovery, achieving optimized compression through dual-domain synergy. Gao et al. [27] employed a spectral decomposition module to separate frequency components, and then utilized a dual-attention mechanism to adaptively fuse the latent features from both domains, achieving improved compression quality. Ma et al.'s iWave++ [88] focuses on a trainable wavelet-like transform, utilizing the multi-scale decomposition capability of wavelet domain to enable lossless conversion.

Image Super-resolution. Image super-resolution is a classic computer vision task dedicated to recovering high-resolution images from low-resolution ones. Despite remarkable progress, conventional spatial-domain deep learning approaches are restricted by lower computational efficiency, inadequate recovery of high-frequency details and poor adaptation to complex real-world scenes. With inherent advantages of frequency analysis including the global perspective, computational efficiency and image disentanglement, FDL offers a promising alternative for tackling complex real-world super-resolution tasks.

In order to capture global topology and local texture details, earlier works [17,37,41,193] leveraged the multiscale decomposition of DWT to isolate low and high frequency components. WaveletSRNet [41] and DWSR [37] frame the image super-resolution problem as wavelet coefficient prediction task. The high-resolution image is then reconstructed from these predicted coefficients via an inverse transform. The reconstruction quality is then enhanced by leveraging complementary information among subbands. Benefit from the multi-resolution property of DWT, SRCliqueNet [193] improves high-frequency detail reconstruction through a network jointly learning four subbands.

 doi:10.20944/preprints202605.0650.v1

5.2. High-Level Vision

5.2.1. Image Classification

In image classification tasks, images are transformed from the spatial domain to the frequency domain in frequency-based deep learning, where frequency characteristics are utilized for feature extraction or computational acceleration. Although CNN-based methods have attained remarkable success in tackling image classification tasks, lots of issues still remain to be addressed. Traditional CNN-based methods process images directly on raw pixels, yet the size and complexity of images in the spatial domain may lead to efficiency degradation. Williams et al. [143] transformed this process into the wavelet domain and processed the wavelet decomposed subband features using CNNs. This approach not only reduces the processing dimensionality, but also improves the classification accuracy by leveraging the frequency decoupling property. To address the problems that CNNs are susceptible to noise interference and suffer from weak robustness, Li et al. [67] replaced the commonly used downsampling operations in CNNs (e.g., max pooling, average pooling and strided convolution) with discrete wavelet transform. Since most noise in an image resides in high-frequency components, they chose to discard the high-frequency components during inference and only extract high-level features from the low-frequency components. Inspired by Li et al.'s work, Zhao et al. [186] designed a wavelet attention mechanism dedicated to the high-frequency components of images, which is used to capture detailed high-frequency information. Qin et al. [105] mathematically proved that traditional global average pooling (GAP) is a special case of frequency-domain feature decomposition, which is equivalent to retaining only the information of the lowest frequency component (i.e., zero frequency or DC component). They further extended the channel attention mechanism via DCT and achieved excellent classification performance on ImageNets.

In addition, a number of frequency-based Transformers [96–98,109,123,165,178] have exhibited promising performance for image classification tasks as well. Table 2 compares the experimental results of various frequency-domain methods on image classification tasks.

Table 2. Comparisons among different FDL paradigms for image classification on ImageNet-1K.

Method	Publication	Resolution	Params (M)	FLOPs (G)	Top-1 (%)	Top-5 (%)
Transformer-based						
GFNet-Ti [109]	Neurips2021	224×224	7	1.3	74.6	92.2
GFNet-XS [109]		224×224	16	2.9	78.6	94.2
GFNet-S [109]		224×224	25	4.5	80.0	94.9
GFNet-B [109]		224×224	43	7.9	80.7	95.1
GFNet-XS [109]		384×384	18	8.4	80.6	95.4
GFNet-S [109]		384×384	28	13.2	81.7	95.8
GFNet-B [109]		384×384	47	23.3	82.1	95.8
SpectFormer-T [98]	WACV2025	224×224	9	1.8	76.8	93.3
SpectFormer-XS [98]		224×224	20	4.0	80.2	94.7
SpectFormer-S [98]		224×224	32	6.6	81.7	95.6
SpectFormer-B [98]		224×224	57	11.5	82.1	95.7
SpectFormer-XS [98]		384×384	21	13.1	82.1	95.7
SpectFormer-S [98]		384×384	33	22.0	83.0	96.3
SpectFormer-B [98]		384×384	57	37.3	82.9	96.1
Wave-ViT-S* [165]	ECCV2022	224×224	22.7	4.7	83.9	96.6
Wave-ViT-S [165]		224×224	19.8	4.3	82.7	96.2
Wave-ViT-B* [165]		224×224	33.5	7.2	84.8	97.1
Wave-ViT-L* [165]		224×224	57.5	14.8	85.5	97.3
SPANet-S [178]	ICCV2023	224×224	29	4.6	83.1	-
SPANet-M [178]		224×224	42	6.8	83.5	-
SPANet-B [178]		224×224	76	12.0	84.0	-
SVT-H-S [97]	NeurIPS2023	224×224	21.7	3.9	83.1	96.3
SVT-H-S* [97]		224×224	22.0	3.9	84.2	96.9
SVT-H-B* [97]		224×224	32.8	6.3	85.2	97.3
SVT-H-L* [97]		224×224	54.0	12.7	85.7	97.5

 doi:10.20944/preprints202605.0650.v1

Table 2. *Cont.*

Method	Publication	Resolution	Params (M)	FLOPs (G)	Top-1 (%)	Top-5 (%)
LITv2-S [96]	NeurIPS2022	224×224	28	3.7	82.0	-
LITv2-M [96]		224×224	49	7.5	83.3	-
LITv2-B [96]		224×224	87	13.2	83.6	-
LITv2-B [96]		384×384	87	39.7	84.7	-
DFFormer-S18 [123]	AAAI2024	224×224	30	3.8	83.2	-
DFFormer-S36 [123]		224×224	46	7.4	84.3	-
DFFormer-M36 [123]		224×224	64	12.7	84.8	-
DFFormer-B36 [123]		224×224	115	22.1	84.8	-
CDFFormer-S18 [123]		224×224	30	3.9	83.1	-
CDFFormer-S36 [123]		224×224	45	7.5	84.2	-
CDFFormer-M36 [123]		224×224	64	12.7	84.8	-
CDFFormer-B36 [123]		224×224	113	22.5	85.0	-
AFNO [34]	ICLR2022	224×224	16	15.3	80.89	95.39
CNN-based						
FcaNet-LF(ResNet-34) [105]	ICCV2021	224×224	21.95	3.68	74.95	92.16
FcaNet-LF(ResNet-50) [105]		224×224	28.07	4.13	78.43	94.15
FcaNet-LF(ResNet-101) [105]		224×224	49.29	7.86	79.46	94.60
FcaNet-LF(ResNet-152) [105]		224×224	66.77	11.60	79.96	94.94
Else						
AFFNet-ET [46]	ICCV2023	256×256	1.4	0.4	73.0	-
AFFNet-ET [46]		256×256	1.6	0.8	77.0	-
AFFNet [46]		256×256	5.5	1.5	79.8	-

5.2.2. Semantic Segmentation

Traditional spatial domain-based segmentation methods face challenges such as high computational complexity and limited generalization capability when dealing with global dependencies, high-resolution images, and out-of-distribution images. Image segmentation leveraging FDL paradigm addresses these issues by converting images into the Fourier domain, and demonstrates unique advantages.

Lo et al. [85] first explored the possibility of performing semantic segmentation in the DCT domain, dropping the traditional step of decompressing images into the RGB format. Zhang et al. [181] proposed performing self-attention only on low-frequency components. Similarly, Shen et al. [114] used the DCT to compress high-resolution binary masks into compact low-dimensional vectors, preserving low-frequency components and discarding high-frequency components with negligible impact, thereby achieving high-quality and low-complexity mask representations.

Yang et al. [164] achieved style transfer between source and target domains by swapping their low-frequency amplitude components, thereby reducing the impact of domain discrepancy on segmentation performance. Chen et al. [9] studied the problem of difficult segmentation for certain pixels from the perspective of spectral aliasing, and proposed DAF and FreqMix to suppress aliasing. Due to the sensitivity of existing semantic segmentation models to frequency information, AFFormer [18] introduces an adaptive frequency filter to capture frequency components that are beneficial for semantic segmentation. Experimental comparisons of different FDL methods for semantic segmentation on ADE20K dataset [194] are presented in Table 3.

 doi:10.20944/preprints202605.0650.v1

Table 3. Comparisons under different FDL paradigms for semantic segmentation on ADE20K.

Method	Publication	mIoU	Params (M)	FLOPS Resolution	FLOPs (G)
Transformer-based					
Wave-ViT-S [165]	ECCV2022	49.6	-	-	-
Wave-ViT-B [165]		51.5	-	-	-
SPANet-S [178]	ICCV2023	45.4	32	512×512	46
SPANet-M [178]		46.2	45	512×512	57
AFFormer-tiny [18]	AAAI2023	38.7	1.6	512×512	2.8
AFFormer-small [18]		40.2	2.3	512×512	3.6
AFFormer-base [18]		41.8	3.0	512×512	4.6
DFFormer-S18 [123]	AAAI2024	45.1	31.7	-	-
CDFFormer-S18 [123]		44.9	31.4	-	-
DFFormer-S36 [123]		47.5	47.2	-	-
CDFFormer-S36 [123]		46.7	46.5	-	-
DFFormer-M36 [123]		47.6	66.4	-	-
CDFFormer-M36 [123]		48.6	65.2	-	-
CNN-based					
FsaNet-LearnPG	ICCV2021	42.24	-	-	-
FsaNet-Dot-R1 [181]		43.04	-	-	-
FsaNet-Lin-R1 [181]		43.05	-	-	-
FsaNet-Dot-R2 [181]		43.53	-	-	-
FsaNet-Lin-R2 [181]		44.10	-	-	-
MLP-based					
Wave-MLP-T [122]	CVPR2022	41.2	19.3	2048×512	131
Wave-MLP-S [122]		44.4	31.2	2048×512	168
Wave-MLP-M [122]		46.8	43.3	2048×512	231
ELSE					
AFFNet-ET [46]	ICCV2023	33.0	2.2	-	-
AFFNet-T [46]		36.9	3.5	-	-
AFFNet [46]		38.4	6.9	-	-

5.2.3. Object Detection

Deep learning in frequency domain has emerged as a crucial technical paradigm to address complex object detection tasks. By transforming images from the spatial domain to the frequency domain, it uncovers discriminative features that are difficult to capture in spatial-domain.

In the task of camouflaged object detection (COD), some studies [15,39,72,81,117,192] incorporate frequency cues into the detectors. Zhong et al. [192] firstly suggested that COD task should break away from the reliance on the single RGB domain and introduce frequency domain information as an additional clue. Liu et al. [81] proposed adding high-frequency components (HFC) to visual prompts for fine-tuning pre-trained models in order to adapt downstream tasks, by stating that high-frequency components possess domain-invariant properties. Lin et al. [72] proposed to address the problems of high-frequency texture interference and discrimination lossy in COD by suppressing high-frequency texture information and adaptively reinforcing important frequency components. The model can easily mine subtle discriminative features under frequency decomposition. He et al. [39] proposed to address COD from a decomposition perspective, performing wavelet-like decomposition on features of different scales, followed by attention mechanism and feature aggregation on the frequency bands with the richest information. Cong et al. [15] proposed a frequency-perception module based on octave convolution to realize online learning of high- and low-frequency features, and used for coarse localization of camouflaged objects. Sun et al. [117] further proposed to enable deep interaction and fusion between spatial and frequency domain features, thereby obtaining more discriminative feature

 doi:10.20944/preprints202605.0650.v1

representations. Experimental comparisons of different FDL methods for object detection (b) and instance segment (m) on COCO [73] are presented in Table 4.

Table 4. Comparisons among different FDL paradigms for object detection (b) and instance segment (m) on COCO.

Method	Publication	AP^b	AP^b_{50}	AP^b_{75}	AP^m	AP^m_{50}	AP^m_{75}
Transformer-based							
SpectFormer-S-FN [98]	WACV2025	46.2	68.1	50.8	42.0	65.2	45.4
SpectFormer-B-FN [98]		46.9	68.8	51.8	42.7	65.9	45.7
LITv2-S [96]	NeurIPS2022	44.9	-	-	40.8	-	-
LITv2-S* [96]		44.7	-	-	40.7	-	-
LITv2-M [96]		46.8	-	-	42.3	-	-
LITv2-M* [96]		46.5	-	-	42.0	-	-
LITv2-B [96]		47.3	-	-	42.6	-	-
LITv2-B* [96]		46.8	-	-	42.3	-	-
Wave-ViT-S [165]	ECCV2022	46.6	68.7	51.2	42.4	65.5	45.8
Wave-ViT-B [165]		47.6	69.1	52.4	43.0	66.4	46.0
SPANet-S [178]	ICCV2023	44.7	65.7	48.8	40.6	62.9	43.8
SPANet-M [178]		45.2	66.3	49.6	41.0	63.5	44.0
SVT-H-S [97]	NeurIPS2023	46.0	68.1	50.4	41.9	65.0	45.1
CNN-based(Resnet50)							
FcaNet-LF [105]	ICCV2021	40.3	61.9	43.9	36.3	58.3	38.6
FcaNet-TS [105]		40.3	62.0	44.1	36.2	58.6	38.1
FcaNet-NAS [105]		40.3	61.9	43.9	36.3	58.3	38.6
Fsa-Dot-R1 [181]	TIP2023	39.9	-	-	36.1	-	-
Fsa-Dot-R2 [181]		39.9	-	-	36.1	-	-
Fsa-Lin-R1 [181]		40.2	-	-	36.3	-	-
Fsa-Lin-R2 [181]		40.1	-	-	36.3	-	-

5.2.4. Image Generation

Traditional generative adversarial networks (GANs) and diffusion models suffer from inherent limitations. First, upsampling operations (e.g., transposed convolution) are prone to introducing high-frequency artifacts, causing a significant deviation between the spectral distribution of generated images and real data [19,49,54,179]. Second, these networks exhibit spectral bias as general CNNs encountered, prioritizing fitting low-frequency signals while losing high-frequency details [26,49,56, 112,163]. Third, in scenarios with limited data, models are prone to overfitting, making it difficult to learn complete spectral distribution characteristics [20,104,141,161].

Durall et al. [19] discovered that upsampling operations in generative models can lead to spectral distortion and proposed adding a spectral regularization term to force the generator to learn spectral distributions that conform to the patterns of natural images. Jung et al. [54] introduced a spectral discriminator to enable the generator to learn the real distributions in the spatial and frequency domains simultaneously. Gal et al. [26] proposed SWAGAN, which integrates DWT into both generator and discriminator for progressive image generation in frequency domain.

FDL paradigm addresses these issues through three core approaches. First, it directly models the spectral distribution of real data, compelling the generator to learn complete frequency characteristics from low to high frequencies. Second, it employs Fourier or Wavelet analysis tools to decompose images, enabling optimization dedicated to different frequency components. Third, it designs specialized frequency-domain loss functions or discriminators to reinforce supervision over spectral matching, thereby improving the visual fidelity of generated images.

 doi:10.20944/preprints202605.0650.v1

5.3. Video Analysis

Compared with the inputs of traditional deep learning models, videos possess higher dimensions, more redundant information, and more complex spatiotemporal correlations. Frequency-domain analysis technique has also demonstrated significant potential in addressing the core challenges of typical video analysis tasks such as segmentation, super-resolution, deblurring and generation.

In unsupervised video object segmentation, Song et al. [151] proposed a Generalizable Fourier Augmentation (GFA) which performs FFT on the intermediate features of Transformer to decompose them into amplitude components encoding scene style information and phase components carrying semantic information. GFA enhances the amplitude components through Gaussian sampling to generate diverse scene style features, thereby alleviating scene shift; meanwhile, it utilizes EMA operator [59] for online updating the phase components, enabling the model to learn domain-invariant features. Inspired by Fnet [62], Pan et al. [95] proposed Wnet, which replaces the self-attention layers in the transformer encoder with 2D DWT to achieve denoised joint representation of audio and video.

The success of Sora [5] has made video generation a research hotspot nowadays. It can not only generate realistic or imaginative videos based on text prompts, but also possess the capability to simulate the laws of the physical world. However, some models [5,71,190] following the Sora paradigm still have some certain limitations in terms of generation hallucinations and computational consumption. In contrast, frequency-domain decomposition can accurately separate key information from redundant components, enabling efficient compression. WF-VAE [69] innovatively incorporates multi-level wavelet transform, decomposing video signals into low-frequency principal energy components and high-frequency detail components. By establishing a energy flow pathway, low-frequency information can directly flow into the latent space rather than the complex backbone network. ConsisID [176] decouples original features into low-frequency global features and high-frequency intrinsic features, and adapts to the DiT [99] through a differentiated injection strategy.

Video deblurring aims to recover clear frames from blurred video clips, but most traditional deep learning methods rely on temporal priors in the spatial domain and lack the utilization of frequency-domain information. Zhu et al. [199] found that blur degradation in videos can be effectively modeled in the frequency domain, and devised a deblurring method that performs Spectral Prior-guided Alignment on adjacent frames with a global-to-detail strategy. Kim et al. [58] proposed a frequency-aware event-based video deblurring method that addresses the modality differences between events and videos in the spectral domain, thereby achieving reliable feature fusion and obtaining a more robust cross-modal feature representation. Qiu et al. [106] utilized DCT to convert compressed video frames into a series of frequency-based patch representations, and achieve deep feature fusion across frequency bands through a unique frequency attention mechanism. Li et al. [66] pointed out that traditional frequency methods have fixed paradigms and coefficients [105,136,175], making them unable to capture complex information in videos. To solve this problem, they proposed MFPI which can effectively aggregate information by operating the spatial- and energy-frequency components. Xu et al. [153] proposed a novel FFT loss to compensate for the lack of high-frequency details caused by over-smoothing.

5.4. Time Series Analysis

A time series is a sequence of data arranged in chronological order, inherently containing underlying patterns such as trends, seasonality, and periodicity. These patterns are often explicitly characterized in frequency domain. For instance, long-term trends correspond to low-frequency components, short-term fluctuations and periodic changes correspond to high-frequency components, while noise typically manifests as irregular high-frequency signals [134]. Traditional time-domain deep learning models (e.g., LSTM [40], Transformer [130], GNN [111]) have been widely used for various time series tasks, including forecasting, classification and anomaly detection [167]. They mainly rely on feature extraction in temporal dimension, but struggle to directly capture frequency-domain information, resulting in limited capability to model complex frequency patterns. Wang et al. [134]

 doi:10.20944/preprints202605.0650.v1

proposed the Multilevel Wavelet Decomposition Network (mWDN) for classification and forecasting tasks, thus achieving effective integration of wavelet transform and time series analysis. Inspired by time-frequency consistency, Zhang et al. [183] proposed a decomposable time series pre-training model where self-supervised signals are derived from the distance between temporal and frequency components, achieving significant performance improvements across various tasks. Yang et al. [160] designed a Temporal-Spectral fusion mechanism and achieved more superior unsupervised time series representation learning.

5.4.1. Forecasting

The core of modeling time series for forecasting in the frequency domain lies in utilizing frequency analysis to capture underlying global dependencies (e.g., trends and periodicity), and to disentangle noise from meaningful historical information. With Fourier transform, Film [196] can eliminate the high-frequency noise and preserve low-frequency patterns. It also adds a low-rank approximation to accelerate computation. Yi et al. [168] proposed a novel Fourier graph neural network (FourierGNN), which performs matrix multiplication in the Fourier space, significantly reducing the computational complexity. CoST [145] captures discriminative seasonal and trend representations separately through contrastive learning in time and frequency domain, respectively. Jiang et al. [50] proposed a novel channel-wise attention mechanism that adaptively models the frequency dependencies between channels based on DCT. Yi et al. [169] proposed FreTS for time series forecasting, fully leveraging the global view and energy compression of frequency-domain MLPs. Yang et al. [162] proposed Fourier Basis Mapping to address the issues of inconsistent starting cycles and inconsistent sequence length in Fourier-based Long-Term Time Series Forecasting methods.

Transformer-based methods have achieved state-of-the-art results in long-term time series forecasting, but suffer from quadratic computation complexity and face challenges in capturing the global perspective. For time series forecasting tasks, Zhou et al. [197] first proposed to implement the attention mechanism in the form of low-rank approximation transformation, leveraging the sparse representation of time series in the frequency domain to reduce computational complexity. Autoformer [146] proposes a FFT based auto-correlation mechanism to replace traditional self-attention, discovering sub-series dependencies based on the series periodicity. FreEformer [177] introduces an additional learnable matrix into the original attention matrix to enhance the diversity of frequency-domain features. In addition, traditional Transformers also suffer from the frequency bias issue in time series forecasting. To address this problem, Fredformer [102] proposes to ensure fair learning of all frequencies by leveraging sub-frequency-independent normalization and intra-sub-frequency attention.

5.4.2. Classification

Frequency-domain deep learning provides a more comprehensive information dimension for classification tasks by exploring the frequency features of time series. mWDN [134] extracts representative features of sub-series from the decomposed results of different levels via the Residual Classification Flow. Yang et al. [160] claimed that the fine-grained features from time-frequency fusion can distinguish similar patterns. Jiang et al. [48] proposed a novel architecture named MH-TFFN by fusing Mamba with hypergraph neural networks, where the model synchronously extracts time-domain and frequency-domain features through a weight-sharing Mamba module. Tian et al. [125] proposed a data augmentation method called FreRA specifically designed for time series classification. By identifying and preserving important semantic components and performing adaptive perturbation on non-important components, the model generates data of diversity retaining the original semantics.

5.4.3. Anomaly Detection

Zhang et al. [180] pointed out the problem that traditional time-domain methods are more likely to detect point anomalies but struggle to identify seasonal anomalies, and further claimed that frequency-domain analysis is more sensitive to seasonal anomalies. Therefore, they suggested simultaneously

 doi:10.20944/preprints202605.0650.v1

model time-domain and frequency-domain features. Meanwhile, RobustTAD [28] and TFAD [180] explore data augmentation methods in the frequency domain to alleviate the problem of labeled data scarcity. Wang et al. [139] innovatively integrated both global frequency features and local frequency features into the condition of Conditional Variational Autoencoder to reconstruct normal data. Chen et al. [8] designed a novel pattern extraction mechanism in the frequency domain, leveraging the sparsity of the frequency domain to enhance the model's efficiency and generalization ability. FreCT [182] detects abnormal patterns in time series by measuring consistency between the time and frequency domains. To solve the problem of time-frequency granularity discrepancy, Nam et al. [93] suggested simultaneously process time-domain and frequency-domain information using nested sliding windows (NS-window), while aligning the information at the data-point granularity. CATCH [147] proposes a frequency-domain patching operation, where each patch corresponds to a frequency band, thereby enhancing the model's ability to capture fine-grained frequency characteristics.

6. Challenges and Future Frontiers

Frequency-principled deep learning (FDL), by integrating classical frequency analysis tools (e.g., Fourier transform) with modern neural networks, exhibits unique advantages in long-range dependency modeling, representation, generalization, robustness and efficiency. It has been widely applied in computer vision, time-series prediction and multimodal analysis. However, it still faces numerous bottlenecks in terms of basic theory, technical implementation and practical application. This section systematically sorts out the core challenges faced by FDL and the future frontiers.

6.1. Challenges and Limitations

6.1.1. Destruction of Causal Structure

One important feature of the Fourier transform is sacrificing temporal locality in order to pursue global frequency information. When a temporal domain signal is converted into a frequency domain signal via the Fourier transform, the causal sequence is compressed into frequency amplitudes, which means that the information about the chronological order is erased. This is not a characteristic of physical processes, but an inevitable consequence of mathematical transformation. In tasks with high demands for temporal logic such as time series analysis, frequency-domain transform usually requires the participation of global data in computation, which may break the causal relationship in the original data. The destruction of such causal structure can lead the model to learn spurious frequency correlations, thereby impairing the reliability of prediction results.

6.1.2. Inadequate Utilization of Frequency-Domain Features

The high- and low-frequency features of data often carry distinct values. For instance, in images, low frequencies correspond to global contours while high frequencies correspond to detailed textures. However, most current models struggle to accurately separate features across different frequency bands and fail to fully exploit the synergistic effects of high- and low-frequency features.

To sum up, how to overcome the limitations in frequency analysis theory, promote the technical synergy with deep neural networks, and forge a new learning paradigm are the key challenges faced by the current FDL paradigm.

6.2. Future Frontiers

6.2.1. Physics-Inspired Frequency Prior Fusion

Incorporating physical-inspired frequency priors into data processing and model design will effectively improve the reliability and practicality of the model. Many recent works [57,64,87,113,135, 152,164] have fully demonstrated the broad development prospects of frequency priors. In domain adaptation and domain generalization tasks, the amplitude spectrum mainly contains low-level statistical information that is prone to change with domain variations; the phase spectrum mainly retains domain-invariant features such as structure and content information [135,152,164]. In low-light image enhancement, the brightness information that plays a critical role is mainly concentrated in

 doi:10.20944/preprints202605.0650.v1

the amplitude spectrum [64,87,113]. In image denoising tasks, the frequency prior is reflected in the fact that noise is mainly distributed in high-frequency bands, while semantic information is mostly concentrated in low-frequency bands [57]. In summary, these frequency priors depend on the target application scenarios. Furthermore, how to utilize this prior knowledge also constitutes a worthwhile research direction. FourierDiff [87] embeds the frequency priors of images into the pre-trained diffusion model, in order to guide the diffusion sampling process. Some works [135,152,164] leverage frequency priors for data augmentation, and [57] converts frequency priors into regularization terms and imposes constraints in the frequency domain. The superior performance demonstrated by these works provides compelling evidence that integrating frequency priors with new learning paradigms is a worthwhile research frontier.

6.2.2. Frequency-Principled Multimodal Large Models

Most current frequency-based deep learning models mainly rely on unimodal information, which limits their applications in complex scenarios. As a unique perspective distinct from the traditional spatio-temporal domain, the frequency domain may reveal some unique correlations and distinctions among different modality representations. Frequency-assisted multimodal models have initially demonstrated tremendous potential. For instance, through DWT decomposition, Zhu et al. [198] found that infrared images have high information entropy in low-frequency sub-bands containing more stable structural information, while RGB images have high information entropy in high-frequency sub-bands containing more detailed content. By processing these sub-bands with different strategies, the complementary characteristics of RGB images and infrared images can be fully exploited. In low-light image enhancement tasks, Xue et al. [159] leveraged the multimodal semantic information of the CLIP model to guide the frequency-domain diffusion process, effectively alleviating the multimodal feature alignment problem caused by image corruption. Therefore, frequency analysis can serve as a basic theoretical tool for cross-modal fusion in some multimodal large models. Future research can focus on exploring the frequency correlations among different modalities and integrating frequency modules into existing multimodal models to fully unleash the power of large models.

6.2.3. Spatial-Time-Frequency Collaborative Modeling

Frequency analysis, as a widely-used tool in image processing and time-series analysis, still suffers from considerable limitations dedicated to high-dimensional tasks including video processing and world model learning. Traditional deep learning methods tend to process information across these dimensions in an isolated or serial manner. However, spatial, temporal and frequency dimensions are not independent but closely correlated. Therefore, exploring efficient spatiotemporal-frequency collaborative modeling holds tremendous research potential. By integrating the temporal trends in the time domain, topological correlations in the spatial domain and periodic characteristics in the frequency domain, FDL is expected to expand to broader and more sophisticated application scenarios such as 4D generation and world models..

6.2.4. Deep Application of Frequency Principle

Numerous studies [108,112,154,156,157,170,191] have demonstrated that frequency principles can exist as an objective law in many deep learning models, i.e., *Deep Neural Networks tend to gradually fit the objective function from low frequencies to high frequencies during the training process.* This is the spectral bias, essentially a low-frequency priority principle. Despite that, most existing deep learning models do not explicitly leverage this property. How to flexibly apply frequency principles to different scenarios and models is an open question worthy of investigation. From the perspective of image classification, Wang et al. [137] analyzed the learning preferences of neural networks and point out that when neural networks perform classification tasks, certain frequency components play a dominant role. Khayatkhoei et al. [56] and Schwarz et al. [112] provided theoretical explanations and solutions regarding the spectral bias problem of GANs in frequency learning. The training process of diffusion models also exhibits the problem of spectral bias. Qian et al. [104] argued that different frequency

 doi:10.20944/preprints202605.0650.v1

components should be processed in a differentiated method following the law of frequency evolution. Yang et al. [163] proposed a learnable wavelet gating mechanism that selectively attends to the proper frequency at different reverse steps to adapt the frequency evolution law.

Besides the performance improvement brought by the frequency principle, it greatly promotes the explainability of deep models. Therefore, frequency principle as a law, has a enormous application potential for developing explainable AI techniques in various fields.

7. Conclusion

In this article, we formally propose the frequency-principled deep learning (FDL) paradigm from the perspective of frequency principle, and present a comprehensive survey of FDL from the basic theory, algorithms and implementations to practical applications. By systematically sorting out the evolution of core technologies of FDL, its key challenges and future frontiers are outlined. Frequency analysis, as a classic theoretical tool for signal processing, frequency-principled deep learning has broken through the limitations of traditional time-domain or spatial-domain modeling, demonstrating unique advantages in capturing periodic patterns, improving generalization, robustness, explainability and efficiency, and modeling long-range dependencies. In this article, we unveil the frequency principle as a inherent law in various deep learning scenarios. To the best of our knowledge, this is the first systematic and in-depth survey focusing on frequency-principled deep learning, which may arouse new but promising learning paradigm of Frequency-principled Foundation Models towards explainable AGI.

References

1. N. Ahmed, T. Natarajan, and K.R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1):90–93, 1974.
2. Simegnew Yihunie Alaba and John E Ball. Wcnn3d: Wavelet convolutional neural network-based 3d object detection for autonomous driving. *Sensors*, 22(18):7010, 2022.
3. Devansh Arpit, Stanisław Jastrzębski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, et al. A closer look at memorization in deep networks. In *ICML*, pages 233–242. PMLR, 2017.
4. E Oran Brigham. *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.
5. Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, et al. Video generation models as world simulators. *OpenAI Blog*, 1(8):1, 2024.
6. Tom Brosch and Roger Tam. Efficient training of convolutional deep belief networks in the frequency domain for application to high-resolution 2d and 3d images. *Neural computation*, 27(1):211–227, 2015.
7. Karol Chęiński and Paweł Wawrzyński. Dct-conv: Coding filters in convolutional networks with discrete cosine transform. In *IJCNN*, pages 1–6. IEEE, 2020.
8. Feiyi Chen, Yingying Zhang, Zhen Qin, Lunting Fan, Renhe Jiang, Yuxuan Liang, Qingsong Wen, and Shuiguang Deng. Learning multi-pattern normalities in the frequency domain for efficient time series anomaly detection. In *ICDE*, pages 747–760. IEEE, 2024.
9. Linwei Chen, Lin Gu, and Ying Fu. When semantic segmentation meets frequency aliasing. *arXiv preprint arXiv:2403.09065*, 2024.
10. Wenlin Chen, James Wilson, Stephen Tyree, Kilian Q Weinberger, and Yixin Chen. Compressing convolutional neural networks in the frequency domain. In *ACM SIGKDD*, pages 1475–1484, 2016.
11. Yuanqi Chen, Ge Li, Cece Jin, Shan Liu, and Thomas Li. Ssd-gan: Measuring the realness in the spatial and spectral domains. In *AAAI*, volume 35, pages 1105–1112, 2021.
12. Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *NeurIPS*, 33:4479–4488, 2020.
13. Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *ICCV*, pages 4641–4650, 2021.
14. Tianyi Chu, Jiafu Chen, Jiakai Sun, Shuobin Lian, Zhizhong Wang, Zhiwen Zuo, Lei Zhao, Wei Xing, and Dongming Lu. Rethinking fast fourier convolution in image inpainting. In *ICCV*, pages 23195–23205, 2023.
15. Runmin Cong, Mengyao Sun, Sanyi Zhang, Xiaofei Zhou, Wei Zhang, and Yao Zhao. Frequency perception network for camouflaged object detection. In *ACM MM*, pages 1179–1189, 2023.

 doi:10.20944/preprints202605.0650.v1

16. Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Image restoration via frequency selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(2):1093–1108, 2023.
17. Xin Deng, Ren Yang, Mai Xu, and Pier Luigi Dragotti. Wavelet domain style transfer for an effective perception-distortion tradeoff in single image super-resolution. In *ICCV*, pages 3076–3085, 2019.
18. Bo Dong, Pichao Wang, and Fan Wang. Head-free lightweight semantic segmentation with linear transformer. In *AAAI*, volume 37, pages 516–524, 2023.
19. Ricard Durall, Margret Keuper, and Janis Keuper. Watch your up-convolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. In *CVPR*, pages 7890–7899, 2020.
20. Tarik Dzanic, Karan Shah, and Freddie Witherden. Fourier spectrum discrepancies in deep network generated images. *NeurIPS*, 33:3022–3032, 2020.
21. Adam Dziedzic, John Paparrizos, Sanjay Krishnan, Aaron Elmore, and Michael Franklin. Band-limited training and inference for convolutional neural networks. In *ICML*, pages 1745–1754. PMLR, 2019.
22. Shahaf E Finder, Roy Amoyal, Eran Treister, and Oren Freifeld. Wavelet convolutions for large receptive fields. In *ECCV*, pages 363–380. Springer, 2024.
23. Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. DW-GAN: A Discrete Wavelet Transform GAN for NonHomogeneous Dehazing. In *CVPRW*, pages 203–212. IEEE.
24. Shin Fujieda, Kohei Takayama, and Toshiya Hachisuka. Wavelet convolutional neural networks for texture classification. *arXiv*, 2017.
25. Shin Fujieda, Kohei Takayama, and Toshiya Hachisuka. Wavelet convolutional neural networks. *arXiv*, 2018.
26. Rinon Gal, Dana Cohen Hochberg, Amit Bermano, and Daniel Cohen-Or. Swagan: A style-based wavelet-driven generative model. *ACM Transactions on Graphics (TOG)*, 40(4):1–11, 2021.
27. Ge Gao, Pei You, Rong Pan, Shunyuan Han, Yuanyuan Zhang, Yuchao Dai, and Hojae Lee. Neural image compression via attentional multi-scale back projection and frequency decomposition. In *ICCV*, pages 14677–14686, 2021.
28. Jingkun Gao, Xiaomin Song, Qingsong Wen, Pichao Wang, Liang Sun, and Huan Xu. Robusttad: Robust time series anomaly detection via decomposition and convolutional neural networks. *arXiv preprint arXiv:2002.09545*, 2020.
29. Ning Gao, Xingyu Jiang, Xiuhui Zhang, and Yue Deng. Efficient frequency-domain image deraining with contrastive regularization. In *ECCV*, pages 240–257. Springer, 2024.
30. Xin Gao, Tianheng Qiu, Xinyu Zhang, Hanlin Bai, Kang Liu, Xuan Huang, Hu Wei, Guoying Zhang, and Huaping Liu. Efficient Multi-Scale Network with Learnable Discrete Wavelet Transform for Blind Motion Deblurring. In *CVPR*, pages 2733–2742. IEEE.
31. Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
32. Julia Grabinski, Steffen Jung, Janis Keuper, and Margret Keuper. Frequencylowcut pooling-plug and play against catastrophic overfitting. In *ECCV*, pages 36–57. Springer, 2022.
33. Lionel Gueguen, Alex Sergeev, Ben Kadlec, Rosanne Liu, and Jason Yosinski. Faster neural networks straight from jpeg. *NeurIPS*, 31, 2018.
34. John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Adaptive fourier neural operators: Efficient token mixers for transformers. *arXiv*, 2021.
35. Jun Guo and Hongyang Chao. Building dual-domain representations for compression artifacts reduction. In *ECCV*, pages 628–644. Springer, 2016.
36. Shi Guo, Hongwei Yong, Xindong Zhang, Jianqi Ma, and Lei Zhang. Spatial-frequency attention for image denoising. *arXiv preprint arXiv:2302.13598*, 2023.
37. Tiantong Guo, Hojjat Seyed Mousavi, Tiep Huu Vu, and Vishal Monga. Deep wavelet prediction for image super-resolution. In *CVPRW*, pages 104–113, 2017.
38. Xin Guo, Xueyang Fu, Man Zhou, Zhen Huang, Jialun Peng, and Zheng-Jun Zha. Exploring fourier prior for single image rain removal. In *IJCAI*, pages 935–941, 2022.
39. Chunming He, Kai Li, Yachao Zhang, Longxiang Tang, Yulun Zhang, Zhenhua Guo, and Xiu Li. Camouflaged object detection with feature decomposition and edge reconstruction. In *CVPR*, pages 22046–22055, 2023.
40. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
41. Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In *ICCV*, pages 1689–1697, 2017.

 doi:10.20944/preprints202605.0650.v1

42. Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Fsdr: Frequency space domain randomization for domain generalization. In *CVPR*, pages 6891–6902, 2021.
43. Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Rda: Robust domain adaptation via fourier adversarial attacking. In *ICCV*, pages 8988–8999, 2021.
44. Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep Fourier-Based Exposure Correction Network with Spatial-Frequency Interaction. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, volume 13679, pages 163–180. Springer Nature Switzerland.
45. Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *ECCV*, pages 163–180. Springer, 2022.
46. Zhipeng Huang, Zhizheng Zhang, Cuiling Lan, Zheng-Jun Zha, Yan Lu, and Baining Guo. Adaptive frequency filters as efficient global token mixers. In *ICCV*, pages 6049–6059, 2023.
47. Bo Jiang, Jinxing Li, Huafeng Li, Ruxian Li, David Zhang, and Guangming Lu. Enhanced frequency fusion network with dynamic hash attention for image denoising. *Information Fusion*, 92:420–434, 2023.
48. Jianjian Jiang, Chuxin Zhuang, Fangyan Lei, Ziwei Chen, Lintao Xiao, Minjing Liang, and Jianfeng Peng. A novel mamba-hypergraph enhanced time-frequency fusion network for multivariate time series classification. *Complex & Intelligent Systems*, 11(9):380, 2025.
49. Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *ICCV*, pages 13919–13929, 2021.
50. Maowei Jiang, Pengyu Zeng, Kai Wang, Huan Liu, Wenbo Chen, and Haoran Liu. Fecam: Frequency enhanced channel attention mechanism for time series forecasting. *Advanced Engineering Informatics*, 58:102158, 2023.
51. Xingyu Jiang, Ning Gao, Hongkun Dou, Xiuhui Zhang, Xiaoqing Zhong, Yue Deng, and Hongjue Li. Global modeling matters: A fast, lightweight and effective baseline for efficient image restoration. *arXiv preprint arXiv:2507.13663*, 2025.
52. Xingyu Jiang, Xiuhui Zhang, Ning Gao, and Yue Deng. When Fast Fourier Transform Meets Transformer for Image Restoration. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, volume 15103, pages 381–402. Springer Nature Switzerland.
53. Xingyu Jiang, Xiuhui Zhang, Ning Gao, and Yue Deng. When fast fourier transform meets transformer for image restoration. In *ECCV*, pages 381–402. Springer, 2024.
54. Steffen Jung and Margret Keuper. Spectral distribution aware image generation. In *AAAI*, volume 35, pages 1734–1742, 2021.
55. Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, pages 8110–8119, 2020.
56. Mahyar Khayatkhoei and Ahmed Elgammal. Spatial frequency bias in convolutional generative adversarial networks. In *AAAI*, volume 36, pages 7152–7159, 2022.
57. Nahyun Kim, Donggon Jang, Sunhyeok Lee, Bomi Kim, and Dae-Shik Kim. Unsupervised image denoising with frequency domain knowledge. *arXiv preprint arXiv:2111.14362*, 2021.
58. Taewoo Kim, Hoonhee Cho, and Kuk-Jin Yoon. Frequency-aware event-based video deblurring for real-world motion blur. In *CVPR*, pages 24966–24976, 2024.
59. Frank Klinker. Exponential moving average versus moving exponential average. *Mathematische Semesterberichte*, 58(1):97–107, 2011.
60. Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, pages 5886–5895, 2023.
61. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *NeurIPS*, 25, 2012.
62. James Lee-Thorp, Joshua Ainslie, Ilya Eckstein, and Santiago Ontanon. Fnet: Mixing tokens with fourier transforms. In *Proceedings of the 2022 Conference of the north American chapter of the Association for Computational Linguistics: human language technologies*, pages 4296–4313, 2022.
63. A Levinskis. Convolutional neural network feature reduction using wavelet transform. *Elektronika ir Elektrotechnika*, 19(3):61–64, 2013.
64. Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*, 2023.

 doi:10.20944/preprints202605.0650.v1

65. Dong Li, Yidi Liu, Xueyang Fu, Senyan Xu, and Zheng-Jun Zha. Fouriermamba: Fourier learning integration with state space models for image deraining. *arXiv*, 2024.
66. Fei Li, Linfeng Zhang, Zikun Liu, Juan Lei, and Zhenbo Li. Multi-frequency representation enhancement with privilege information for video super-resolution. In *ICCV*, pages 12814–12825, 2023.
67. Qiufu Li, Linlin Shen, Sheng Guo, and Zhihui Lai. Wavelet integrated cnns for noise-robust image classification. In *CVPR*, pages 7245–7254, 2020.
68. Shaohua Li, Kaiping Xue, Bin Zhu, Chenkai Ding, Xindi Gao, David Wei, and Tao Wan. FALCON: A Fourier Transform Based Approach for Fast and Secure Convolutional Neural Network Predictions. In *CVPR*, pages 8702–8711. IEEE.
69. Zongjian Li, Bin Lin, Yang Ye, Liuhan Chen, Xinhua Cheng, Shenghai Yuan, and Li Yuan. Wf-vae: Enhancing video vae by wavelet-driven energy flow for latent video diffusion model. In *CVPR*, pages 17778–17788, 2025.
70. Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.
71. Bin Lin, Yunyang Ge, Xinhua Cheng, Zongjian Li, Bin Zhu, Shaodong Wang, Xianyi He, Yang Ye, Shenghai Yuan, Liuhan Chen, et al. Open-sora plan: Open-source large video generation model. *arXiv preprint arXiv:2412.00131*, 2024.
72. Jiaying Lin, Xin Tan, Ke Xu, Lizhuang Ma, and Rynson WH Lau. Frequency-aware camouflaged object detection. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2):1–16, 2023.
73. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755. Springer, 2014.
74. Chengxu Liu, Lu Qi, Jinshan Pan, Xueming Qian, and Ming-Hsuan Yang. Frequency domain-based diffusion model for unpaired image dehazing. *arXiv preprint arXiv:2507.01275*, 2025.
75. Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. Pay attention to mlps. *NeurIPS*, 34:9204–9215, 2021.
76. Lin Liu, Jianzhuang Liu, Shanxin Yuan, Gregory Slabaugh, Aleš Leonardis, Wengang Zhou, and Qi Tian. Wavelet-based dual-branch network for image demoiréing. In *ECCV*, pages 86–102. Springer, 2020.
77. Pengju Liu, Hongzhi Zhang, Wei Lian, and Wangmeng Zuo. Multi-level wavelet convolutional neural networks. *IEEE Access*, 7:74973–74985, 2019.
78. Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *CVPRW*, pages 773–782, 2018.
79. Quande Liu, Cheng Chen, Jing Qin, Qi Dou, and Pheng-Ann Heng. FedDG: Federated Domain Generalization on Medical Image Segmentation via Episodic Learning in Continuous Frequency Space. In *CVPR*, pages 1013–1023. IEEE.
80. Weihuang Liu, Xi Shen, Chi-Man Pun, and Xiaodong Cun. Explicit Visual Prompting for Low-Level Structure Segmentations. In *CVPR*, pages 19434–19445.
81. Weihuang Liu, Xi Shen, Chi-Man Pun, and Xiaodong Cun. Explicit visual prompting for low-level structure segmentations. In *CVPR*, pages 19434–19445, 2023.
82. Xiao Liu, Chenxu Zhang, Fuxiang Huang, Shuyin Xia, Guoyin Wang, and Lei Zhang. Vision mamba: A comprehensive survey and taxonomy. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.
83. Yunfan Liu, Qi Li, and Zhenan Sun. Attribute-aware face aging with wavelet-based generative adversarial networks. In *CVPR*, pages 11877–11886, 2019.
84. Zhenhua Liu, Jizheng Xu, Xiulian Peng, and Ruiqin Xiong. Frequency-domain dynamic pruning for convolutional neural networks. *NeurIPS*, 31, 2018.
85. Shao-Yuan Lo and Hsueh-Ming Hang. Exploring semantic segmentation on the dct representation. In *Proceedings of the 1st ACM International Conference on Multimedia in Asia*, pages 1–6, 2019.
86. Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier Priors-Guided Diffusion for Zero-Shot Joint Low-Light Enhancement and Deblurring. In *CVPR*, pages 25378–25388. IEEE.
87. Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *CVPR*, pages 25378–25388, 2024.
88. Haichuan Ma, Dong Liu, Ning Yan, Houqiang Li, and Feng Wu. End-to-end optimized versatile image compression with wavelet-like transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3):1247–1263, 2020.

 doi:10.20944/preprints202605.0650.v1

89. Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 2002.
90. Xintian Mao, Yiming Liu, Wei Shen, Qingli Li, and Yan Wang. Deep residual fourier transformation for single image deblurring. *arXiv preprint arXiv:2111.11745*, 2(3):5, 2021.
91. Michael Mathieu, Mikael Henaff, and Yann LeCun. Fast training of convolutional networks through ffts. *arXiv*, 2013.
92. Wei Miao, Jiangrong Shen, Qi Xu, Timo Hamalainen, Yi Xu, and Fengyu Cong. Spikingyolox: Improved yolox object detection with fast fourier convolution and spiking neural networks. In *AAAI*, volume 39, pages 1465–1473, 2025.
93. Youngeun Nam, Susik Yoon, Yooju Shin, Minyoung Bae, Hwanjun Song, Jae-Gil Lee, and Byung Suk Lee. Breaking the time-frequency granularity discrepancy in time-series anomaly detection. In *Proceedings of the ACM Web Conference 2024*, pages 4204–4215, 2024.
94. Mang Ning, Mingxiao Li, Jianlin Su, Haozhe Jia, Lanmiao Liu, Martin Beneš, Wenshuo Chen, Albert Ali Salah, and Itir Onal Ertugrul. Dctdiff: Intriguing properties of image generative modeling in the dct space. *arXiv preprint arXiv:2412.15032*, 2024.
95. Wenwen Pan, Haonan Shi, Zhou Zhao, Jieming Zhu, Xiuqiang He, Zhigeng Pan, Lianli Gao, Jun Yu, Fei Wu, and Qi Tian. Wnet: Audio-guided video object segmentation via wavelet-based cross-modal denoising networks. In *CVPR*, pages 1320–1331, 2022.
96. Zizheng Pan, Jianfei Cai, and Bohan Zhuang. Fast vision transformers with hilo attention. *NeurIPS*, 35:14541–14554, 2022.
97. Badri Patro and Vijay Agneeswaran. Scattering vision transformer: Spectral mixing matters. *NeurIPS*, 36:54152–54166, 2023.
98. Badri N Patro, Vinay P Namboodiri, and Vijay S Agneeswaran. Spectformer: Frequency and attention is what you need in a vision transformer. In *WACV*, pages 9543–9554, 2025.
99. William Peebles and Saining Xie. Scalable diffusion models with transformers. In *ICCV*, pages 4195–4205, 2023.
100. Cuong Pham, Van-Anh Nguyen, Trung Le, Dinh Phung, Gustavo Carneiro, and Thanh-Toan Do. Frequency Attention for Knowledge Distillation. In *WACV*, pages 2266–2275. IEEE.
101. Hao Phung, Quan Dao, and Anh Tran. Wavelet diffusion models are fast and scalable image generators. In *CVPR*, pages 10199–10208, 2023.
102. Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *ACM SIGKDD*, pages 2400–2410, 2024.
103. Simeon Denis Poisson. Mémoire sur la propagation de la chaleur dans les corps solides. *Nouveau Bulletin des Sciences par la Société philomatique de Paris, tI*, pages 112–116, 1808.
104. Yurui Qian, Qi Cai, Yingwei Pan, Yehao Li, Ting Yao, Qibin Sun, and Tao Mei. Boosting diffusion models with moving average sampling in frequency domain. In *CVPR*, pages 8911–8920, 2024.
105. Zequn Qin, Pengyi Zhang, Fei Wu, and Xi Li. Fcanet: Frequency channel attention networks. In *ICCV*, pages 783–792, 2021.
106. Zhongwei Qiu, Huan Yang, Jianlong Fu, and Dongmei Fu. Learning spatiotemporal frequency-transformer for compressed video super-resolution. In *ECCV*, pages 257–273. Springer, 2022.
107. Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
108. Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *ICML*, pages 5301–5310. PMLR, 2019.
109. Yongming Rao, Wenliang Zhao, Zheng Zhu, Jiwen Lu, and Jie Zhou. Global filter networks for image classification. *NeurIPS*, 34:980–993, 2021.
110. Oren Rippel, Jasper Snoek, and Ryan P Adams. Spectral representations for convolutional neural networks. *NeurIPS*, 28, 2015.
111. Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.
112. Katja Schwarz, Yiyi Liao, and Andreas Geiger. On the frequency bias of generative models. *NeurIPS*, 34:18126–18136, 2021.
113. Chunyan She, Fujun Han, Chengyu Fang, Shukai Duan, and Lidan Wang. Exploring fourier prior and event collaboration for low-light image enhancement. In *ACM MM*, pages 3017–3026, 2025.

 doi:10.20944/preprints202605.0650.v1

114. Xing Shen, Jirui Yang, Chunbo Wei, Bing Deng, Jianqiang Huang, Xian-Sheng Hua, Xiaoliang Cheng, and Kewei Liang. Dct-mask: Discrete cosine transform mask representation for instance segmentation. In *CVPR*, pages 8720–8729, 2021.
115. Zehua Sheng, Xiongwei Liu, Si-Yuan Cao, Hui-Liang Shen, and Huaqi Zhang. Frequency-domain deep guided image denoising. *IEEE Transactions on Multimedia*, 25:6767–6781, 2022.
116. Samir S Soliman and Mandyam D Srinath. *Continuous and discrete signals and systems*. Prentice-Hall, Inc., 1990.
117. Yanguang Sun, Chunyan Xu, Jian Yang, Hanyu Xuan, and Lei Luo. Frequency-spatial entanglement learning for camouflaged object detection. In *ECCV*, pages 343–360. Springer, 2024.
118. Duraisamy Sundararajan. *The discrete Fourier transform: theory, algorithms and applications*. World Scientific, 2001.
119. Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust Large Mask Inpainting with Fourier Convolutions. In *WACV*, pages 3172–3182. IEEE.
120. Junhao Tan, Songwen Pei, Wei Qin, Bo Fu, Ximing Li, and Libo Huang. Wavelet-based mamba with fourier adjustment for low-light image enhancement. In *ACCV*, pages 3449–3464, 2024.
121. Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 33:7537–7547, 2020.
122. Yehui Tang, Kai Han, Jianyuan Guo, Chang Xu, Yanxi Li, Chao Xu, and Yunhe Wang. An image patch is a wave: Phase-aware vision mlp. In *CVPR*, pages 10935–10944, 2022.
123. Yuki Tatsunami and Masato Taki. Fft-based dynamic token mixer for vision. In *AAAI*, volume 38, pages 15328–15336, 2024.
124. Chunwei Tian, Menghua Zheng, Wangmeng Zuo, Bob Zhang, Yanning Zhang, and David Zhang. Multi-stage image denoising with the wavelet transform. *Pattern Recognition*, 134:109050, 2023.
125. Tian Tian, Chunyan Miao, and Hangwei Qian. Frera: a frequency-refined augmentation for contrastive learning on time series classification. In *ACM SIGKDD*, pages 2835–2846, 2025.
126. Ilya O Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, et al. Mlp-mixer: An all-mlp architecture for vision. *NeurIPS*, 34:24261–24272, 2021.
127. Hugo Touvron, Piotr Bojanowski, Mathilde Caron, Matthieu Cord, Alaaeldin El-Nouby, Edouard Grave, Gautier Izacard, Armand Joulin, Gabriel Synnaeve, Jakob Verbeek, et al. Resmlp: Feedforward networks for image classification with data-efficient training. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):5314–5321, 2022.
128. Matej Ulicny, Vladimir A Krylov, and Rozenn Dahyot. Harmonic convolutional networks based on discrete cosine transform. *Pattern Recognition*, 129:108707, 2022.
129. Nicolas Vasilache, Jeff Johnson, Michael Mathieu, Soumith Chintala, Serkan Piantino, and Yann LeCun. Fast convolutional nets with fbfft: A gpu performance evaluation. *arXiv*, 2014.
130. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 30, 2017.
131. Gregory K Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991.
132. Chenyang Wang, Junjun Jiang, Zhiwei Zhong, and Xianming Liu. Spatial-frequency mutual learning for face super-resolution. In *CVPR*, pages 22356–22366, 2023.
133. Jingye Wang, Ruoyi Du, Dongliang Chang, Kongming Liang, and Zhanyu Ma. Domain generalization via frequency-domain-based feature disentanglement and interaction. In *ACM MM*, pages 4821–4829, 2022.
134. Jingyuan Wang, Ze Wang, Jianfeng Li, and Junjie Wu. Multilevel wavelet decomposition network for interpretable time series analysis. In *ACM SIGKDD*, pages 2437–2446, 2018.
135. Kunyu Wang, Xueyang Fu, Yukun Huang, Chengzhi Cao, Gege Shi, and Zheng-Jun Zha. Generalized uav object detection via frequency domain disentanglement. In *CVPR*, pages 1064–1073, 2023.
136. Peihao Wang, Wenqing Zheng, Tianlong Chen, and Zhangyang Wang. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. *arXiv*, 2022.
137. Shunxin Wang, Raymond Veldhuis, Christoph Brune, and Nicola Strisciuglio. What do neural networks learn in image classification? a frequency shortcut perspective. In *ICCV*, pages 1433–1442, 2023.

 doi:10.20944/preprints202605.0650.v1

138. Yunhe Wang, Chang Xu, Chao Xu, and Dacheng Tao. Packing convolutional neural networks in the frequency domain. *IEEE transactions on pattern analysis and machine intelligence*, 41(10):2495–2510, 2018.
139. Zexin Wang, Changhua Pei, Minghua Ma, Xin Wang, Zhihan Li, Dan Pei, Saravan Rajmohan, Dongmei Zhang, Qingwei Lin, Haiming Zhang, et al. Revisiting vae for unsupervised time series anomaly detection: A frequency perspective. In *Proceedings of the ACM web conference 2024*, pages 3096–3105, 2024.
140. Zhangyang Wang, Ding Liu, Shiyu Chang, Qing Ling, Yingzhen Yang, and Thomas S Huang. D3: Deep dual-domain based fast restoration of jpeg-compressed images. In *CVPR*, pages 2764–2772, 2016.
141. Zhe Wang, Ziqiu Chi, Yanbing Zhang, et al. Fregan: Exploiting frequency components for training gans under limited data. *NeurIPS*, 35:33387–33399, 2022.
142. Travis Williams and Robert Li. WAVELET POOLING FOR CONVOLUTIONAL NEURAL NETWORKS.
143. Travis Williams and Robert Li. Advanced image classification using wavelets and convolutional neural networks. In *ICMLA*, pages 233–239. IEEE, 2016.
144. Travis Williams and Robert Li. Wavelet pooling for convolutional neural networks. In *ICLR*, 2018.
145. Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. *arXiv*, 2022.
146. Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *NeurIPS*, 34:22419–22430, 2021.
147. Xingjian Wu, Xiangfei Qiu, Zhengyu Li, Yihang Wang, Jilin Hu, Chenjuan Guo, Hui Xiong, and Bin Yang. Catch: Channel-aware multivariate time series anomaly detection via frequency patching. *arXiv preprint arXiv:2410.12261*, 2024.
148. Yi Xiao, Qiangqiang Yuan, Kui Jiang, Yuzeng Chen, Qiang Zhang, and Chia-Wen Lin. Frequency-assisted mamba for remote sensing image super-resolution. *IEEE Transactions on Multimedia*, 2024.
149. Zipeng Xiao, Siqi Kou, Zhongkai Hao, Bokai Lin, and Zhijie Deng. Amortized fourier neural operators. *NeurIPS*, 37:115001–115020, 2024.
150. Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain. In *CVPR*, pages 1740–1749, 2020.
151. Mingjun Xu, Lingyun Qin, Weijie Chen, Shiliang Pu, and Lei Zhang. Multi-view adversarial discriminator: Mine the non-causal factors for object detection in unseen domains. In *CVPR*, pages 8103–8112, 2023.
152. Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *CVPR*, pages 14383–14392, 2021.
153. Yi Xu, Minyi Zhao, Jing Liu, Xinjian Zhang, Longwen Gao, Shuigeng Zhou, and Huyang Sun. Boosting the performance of video compression artifact reduction with reference frame proposals and frequency domain information. In *CVPR*, pages 213–222, 2021.
154. Zhi-Qin John Xu. Frequency principle in deep learning with general loss functions and its potential application. *arXiv preprint arXiv:1811.10146*, 2018.
155. Zhi-Qin John Xu, Yaoyu Zhang, and Tao Luo. Overview frequency principle/spectral bias in deep learning. *Communications on Applied Mathematics and Computation*, 7(3):827–864, 2025.
156. Zhi-Qin John Xu, Yaoyu Zhang, and Yanyang Xiao. Training behavior of deep neural network in frequency domain. In *ICONIP*, pages 264–274. Springer, 2019.
157. Zhiqin John Xu. Understanding training and generalization in deep learning by fourier analysis. *arXiv*, 2018.
158. Zhiqin John Xu and Hanxu Zhou. Deep frequency principle towards understanding why deeper learning is faster. In *AAAI*, volume 35, pages 10541–10550, 2021.
159. Minglong Xue, Jinhong He, Wenhai Wang, and Mingliang Zhou. Low-light image enhancement via clip-fourier guided wavelet diffusion. *ACM Transactions on Multimedia Computing, Communications and Applications*, 21(11):1–22, 2025.
160. Ling Yang and Shenda Hong. Unsupervised time-series representation learning with iterative bilinear temporal-spectral fusion. In *ICML*, pages 25038–25054. PMLR, 2022.
161. Mengping Yang, Zhe Wang, Ziqiu Chi, and Wenyi Feng. Wavegan: Frequency-aware gan for high-fidelity few-shot image generation. In *ECCV*, pages 1–17. Springer, 2022.
162. Runze Yang, Longbing Cao, JIE YANG, et al. Rethinking fourier transform from a basis functions perspective for long-term time series forecasting. *NeurIPS*, 37:8515–8540, 2024.
163. Xingyi Yang, Daquan Zhou, Jiashi Feng, and Xinchao Wang. Diffusion probabilistic model made slim. In *CVPR*, pages 22552–22562, 2023.

 doi:10.20944/preprints202605.0650.v1

164. Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *CVPR*, pages 4085–4095, 2020.
165. Ting Yao, Yingwei Pan, Yehao Li, Chong-Wah Ngo, and Tao Mei. Wave-vit: Unifying wavelet and transformers for visual representation learning. In *ECCV*, pages 328–345. Springer, 2022.
166. Zishu Yao, Guodong Fan, Jinfu Fan, Min Gan, and CL Philip Chen. Spatial-frequency dual-domain feature fusion network for low-light remote sensing image enhancement. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.
167. Kun Yi, Qi Zhang, Wei Fan, Longbing Cao, Shoujin Wang, Hui He, Guodong Long, Liang Hu, Qingsong Wen, and Hui Xiong. A survey on deep learning based time series analysis with frequency transformation. In *ACM SIGKDD*, pages 6206–6215, 2025.
168. Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective. *NeurIPS*, 36:69638–69660, 2023.
169. Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *NeurIPS*, 36:76656–76679, 2023.
170. Dong Yin, Raphael Gontijo Lopes, Jon Shlens, Ekin Dogus Cubuk, and Justin Gilmer. A fourier perspective on model robustness in computer vision. *NeurIPS*, 32, 2019.
171. Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and Spatial Dual Guidance for Image Dehazing. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, volume 13679, pages 181–198. Springer Nature Switzerland.
172. Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In *ECCV*, pages 181–198. Springer, 2022.
173. Weihao Yu, Mi Luo, Pan Zhou, Chenyang Si, Yichen Zhou, Xinchao Wang, Jiashi Feng, and Shuicheng Yan. Metaformer is actually what you need for vision. In *CVPR*, pages 10819–10829, June 2022.
174. Yingchen Yu, Fangneng Zhan, Shijian Lu, Jianxiong Pan, Feiying Ma, Xuansong Xie, and Chunyan Miao. WaveFill: A Wavelet-based Generation Network for Image Inpainting. In *ICCV*, pages 14094–14103. IEEE.
175. Yingchen Yu, Fangneng Zhan, Shijian Lu, Jianxiong Pan, Feiying Ma, Xuansong Xie, and Chunyan Miao. Wavefill: A wavelet-based generation network for image inpainting. In *ICCV*, pages 14114–14123, 2021.
176. Shenghai Yuan, Jinfa Huang, Xianyi He, Yunyang Ge, Yujun Shi, Liuhan Chen, Jiebo Luo, and Li Yuan. Identity-preserving text-to-video generation by frequency decomposition. In *CVPR*, pages 12978–12988, 2025.
177. Wenzhen Yue, Yong Liu, Xianghua Ying, Bowei Xing, Ruohao Guo, and Ji Shi. Freeformer: Frequency enhanced transformer for multivariate time series forecasting. *arXiv*, 2025.
178. Guhnoo Yun, Juhan Yoo, Kijung Kim, Jeongho Lee, and Dong Hwan Kim. Spanet: Frequency-balancing token mixer using spectral pooling aggregation modulation. In *ICCV*, pages 6113–6124, 2023.
179. Bowen Zhang, Shuyang Gu, Bo Zhang, Jianmin Bao, Dong Chen, Fang Wen, Yong Wang, and Baining Guo. Styleswin: Transformer-based gan for high-resolution image generation. In *CVPR*, pages 11304–11314, 2022.
180. Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. Tfad: A decomposition time series anomaly detection architecture with time-frequency analysis. In *ACM CIKM*, pages 2497–2507, 2022.
181. Fengyu Zhang, Ashkan Panahi, and Guangjun Gao. Fsanet: Frequency self-attention for semantic segmentation. *IEEE Transactions on Image Processing*, 32:4757–4772, 2023.
182. Wenxin Zhang, Ding Xu, Guangzhen Yao, Xiaojian Lin, Renxiang Guan, Chengze Du, Renda Han, Xi Xuan, and Cuicui Luo. Frect: Frequency-augmented convolutional transformer for robust time series anomaly detection. In *ICIC*, pages 15–26, 2025.
183. Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. Self-supervised contrastive pre-training for time series via time-frequency consistency. *NeurIPS*, 35:3988–4003, 2022.
184. Zhendong Zhang, Cheolkon Jung, and Xiaolong Liang. Adversarial defense by suppressing high-frequency components. *arXiv preprint arXiv:1908.06566*, 2019.
185. Chen Zhao, Weiling Cai, Chenyu Dong, and Chengwei Hu. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *CVPR*, pages 8281–8291, 2024.
186. Xiangyu Zhao, Peng Huang, and Xiangbo Shu. Wavelet-attention cnn for image classification. *Multimedia Systems*, 28(3):915–924, 2022.

 doi:10.20944/preprints202605.0650.v1

187. Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Ales Leonardis. Image demoireing with learnable bandpass filters. In *CVPR*, pages 3636–3645, 2020.
188. Bolun Zheng, Shanxin Yuan, Chenggang Yan, Xiang Tian, Jiyong Zhang, Yaoqi Sun, Lin Liu, Ales Leonardis, and Gregory Slabaugh. Learning Frequency Domain Priors for Image Demoireing. 44(11):7705–7717.
189. Bolun Zheng, Shanxin Yuan, Chenggang Yan, Xiang Tian, Jiyong Zhang, Yaoqi Sun, Lin Liu, Aleš Leonardis, and Gregory Slabaugh. Learning frequency domain priors for image demoireing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7705–7717, 2021.
190. Zangwei Zheng, Xiangyu Peng, Tianji Yang, Chenhui Shen, Shenggui Li, Hongxin Liu, Yukun Zhou, Tianyi Li, and Yang You. Open-sora: Democratizing efficient video production for all. *arXiv preprint arXiv:2412.20404*, 2024.
191. John Zhi-Qin, John Xu, Luo Tao, Xiao Yanyang, and Ma Zheng. Frequency principle: Fourier analysis sheds light on deep neural networks. *Communications in Computational Physics*, 28(5):1746–1767, 2020.
192. Yijie Zhong, Bo Li, Lv Tang, Senyun Kuang, Shuang Wu, and Shouhong Ding. Detecting camouflaged object in frequency domain. In *CVPR*, pages 4504–4513, 2022.
193. Zhisheng Zhong, Tiancheng Shen, Yibo Yang, Zhouchen Lin, and Chao Zhang. Joint sub-bands learning with clique structures for wavelet domain super-resolution. *NeurIPS*, 31, 2018.
194. Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *CVPR*, July 2017.
195. Man Zhou, Jie Huang, Chun-Le Guo, and Chongyi Li. Fourmer: An efficient global modeling paradigm for image restoration. In *ICML*, pages 42589–42601. PMLR, 2023.
196. Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al. Film: Frequency improved legendre memory model for long-term time series forecasting. *NeurIPS*, 35:12677–12690, 2022.
197. Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *ICML*, pages 27268–27286. PMLR, 2022.
198. Haodong Zhu, Wenhao Dong, Linlin Yang, Hong Li, Yuguang Yang, Yangyang Ren, Qingcheng Zhu, Zichao Feng, Changbai Li, Shaohui Lin, et al. Wavemamba: Wavelet-driven mamba fusion for rgb-infrared object detection. In *ICCV*, pages 11219–11229, 2025.
199. Qi Zhu, Man Zhou, Naishan Zheng, Chongyi Li, Jie Huang, and Feng Zhao. Exploring temporal frequency spectrum in deep video deblurring. In *ICCV*, pages 12428–12437, 2023.
200. Wenbin Zou, Hongxia Gao, Weipeng Yang, and Tongtong Liu. Wave-mamba: Wavelet state space model for ultra-high-definition low-light image enhancement. In *ACM MM*, pages 1534–1543, 2024.
201. Wenbin Zou, Mingchao Jiang, Yunchen Zhang, Liang Chen, Zhiyong Lu, and Yi Wu. SDWNet: A Straight Dilated Network with Wavelet Transformation for image Deblurring. In *ICCVW*, pages 1895–1904. IEEE.
202. Zhen Zou, Hu Yu, Jie Huang, and Feng Zhao. Freqmamba: Viewing mamba from a frequency perspective for image deraining. In *ACM MM*, pages 1905–1914, 2024.